



TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA

**Análisis del valor pronóstico de los
TRECs (*T-cell receptor excision
circles*) y KRECs (*K-deleting
recombination excision circles*) en
resultados de trasplante alogénico
mediante modelos *joint***

Julio 2020

David López Domínguez

Tutor: Pedro Girón Daviña

Tutora: Ana Royuela Vicente

INDICE

RESUMEN	4
ABSTRACT	5
ÍNDICE DE ABREVIATURAS	6
1. INTRODUCCIÓN	7
2. HIPÓTESIS Y OBJETIVOS.....	11
3. METODOLOGÍA.....	12
3.1. DISEÑO DEL ESTUDIO Y MATERIAL.....	12
3.1.1 Tipo de estudio, población y base de datos.....	12
3.1.2 Desenlaces clínicos principales y secundarios	14
3.2. METODOLOGÍA ESTADÍSTICA	15
3.2.1 Análisis descriptivo. Imputaciones y transformación de TRECs y KRECs.	15
3.2.2 Modelos Mixtos para el análisis de datos longitudinales	16
3.2.3 Análisis de supervivencia para tiempo hasta evento.....	18
3.2.4 Análisis de supervivencia con eventos competitivos.	20
3.2.5 Modelos conjuntos de verosimilitud (modelos <i>joint</i>)	21
4 RESULTADOS	27
4.1 Análisis descriptivo.....	27
4.2 Modelos de efectos mixtos para cada respuesta longitudinal.	33
4.3 Modelos de tiempo hasta el evento para cada biomarcador.....	39
4.4 Modelos <i>joint</i> para cada biomarcador	41
4.5 Modelos <i>joint</i> para respuesta longitudinal multivariante.....	45
4.6 Modelos <i>joint</i> para riesgos competitivos con modelos multiestado	48
5 FORTALEZAS Y LIMITACIONES	56
6 CONCLUSIONES	58
7 BIBLIOGRAFÍA.....	60
ANEXO	63

RESUMEN

En el campo de la biomedicina, es habitual recoger información periódicamente de variables endógenas como son los biomarcadores, útiles como indicadores del estado biológico del paciente. Como ejemplo de estos biomarcadores están los TRECs (*T-cell receptor excision circles*) y KRECs (*K-deleting excision circles*), que permiten cuantificar, en cierta medida, el grado de reconstitución inmune adaptativa mediada tanto por linfocitos T (TRECs) como B (KRECs) en pacientes con patología hematológica.

El objetivo del estudio es analizar la asociación entre la respuesta longitudinal de los biomarcadores TRECs y KRECs con el tiempo libre de progresión de la enfermedad (PFS), así como con los eventos competitivos recaída y fallecimiento sin recaída (NRM) a los dos años del trasplante alogénico. Para ello, se considera una cohorte de pacientes diagnosticados con una patología hematológica determinada, sometidos a trasplante de células madre hematopoyéticas (HSCT) aplicando la metodología de los modelos *joint*, tanto de forma univariante para cada biomarcador por separado, como de forma multivariante ajustando ambos, así como abordando un análisis de riesgos competitivos mediante un enfoque multiestado sin estados intermedios.

Los resultados obtenidos indican una asociación protectora tanto de los TRECs como de los KRECs con el riesgo sobre PFS a los dos años del trasplante, si bien la asociación de los KRECs parece quedar recogida o explicada por los TRECs cuando se ajustan ambas variables en un modelo multivariante. En el análisis de riesgos competitivos se mantiene la misma tendencia al realizar las estimaciones univariantes, determinándose ambos como factores protectores para el fallecimiento sin recaída, y únicamente los niveles de TRECs sobre el tiempo hasta recaída. Al ajustar ambos biomarcadores, los TRECs mantienen la asociación con el riesgo de recaída, mientras que los KRECs recogen el efecto de los TRECs sobre el tiempo hasta fallecimiento sin recaída.

Palabras clave: TRECs, KRECs, biomarcador, trasplante HSCT, Modelo *joint*, PFS.

ABSTRACT

In biomedicine, it is common to collect regularly information about endogenous variables such as biomarkers, useful as indicators of the biological status of a patient. TRECs (*T-cell receptor excision circles*) and KRECs (*K-deleting excision circles*) are an example of these biomarkers, which allow to measure the degree of reconstitution of the adaptive immunity, carried out by both T (TRECs) and B (KRECs) cells in patients with an hematological disease.

The aim in the present study is to ascertain the association between the longitudinal response of TRECs and KRECs with the progression free survival time (PFS), as well as with the competitive events of relapse and non-relapse mortality (NRM), in the 2 years posterior to an allogenic bone marrow transplantation. To this end, a cohort of patients diagnosed with an hematological disease and which underwent hematopoietic stem cell transplantation (HSCT) was evaluated. The joint modelling approach was performed, both by univariate and multivariate analysis, as well as a competitive risk analysis, carried out by a multistate approach without intermediate events.

The obtained results indicate a protective association between both TRECs and KRECs with PFS at 2 years from the HSCT, although the KRECs association seems to be explained by TRECs when fitting both biomarkers in a multivariate model. That tendency is maintained after performing the univariate competitive risk analysis, showing both biomarkers as protective factors for NRM, but only TRECs levels for relapse survival time. When fitting both biomarkers, TRECs association with relapse is maintained, while KRECs reflects the TRECs univariate effect over NRM survival time.

Key words: TRECs, KRECs, biomarker, HSCT transplantation, joint models, PFS.

ÍNDICE DE ABREVIATURAS

HSCT: Trasplante de células madre hematopoyéticas (*Hematopoietic Stem Cell Transplantation*).

KRECs: Círculos de escisión de la cadena Kappa (*K-deleting Recombination Excision Circles*).

LOCF: Última observación arrastrada hacia delante (*Last Observation Carried Forward*)

NRM: Fallecimiento sin recaída (*Non-Relapse Mortality*)

MCMC: Método de cadenas de Markov Monte Carlo (*Monte Carlo Markov Chain*)

PCR: Reacción en cadena de la polimerasa (*Polymerase Chain Reaction*)

PFS: Supervivencia libre de progresión (*Progression Free Survival*)

TCR: Receptor de células T (*T-Cell Receptor*)

TPH: Trasplante alogénico

TRECs: Círculos de escisión del receptor de células T (*T-cell Receptor Excision Circles*).

1. INTRODUCCIÓN

En el ámbito clínico, se ha considerado fundamental el poder evaluar la evolución de variables a lo largo del tiempo, con el objetivo no sólo de conocer su comportamiento, sino también de poder determinar la asociación de dicha evolución con el pronóstico para pacientes u otras unidades experimentales ^[1]. El análisis de estos datos, denominados datos longitudinales, implica la medición a lo largo del tiempo de características clínicas de un mismo paciente o respuestas a determinados procesos exógenos, permitiendo estimar la relación existente entre su evolución con la de otras variables específicas.

En dicho ámbito, los pacientes que padecen una patología determinada son generalmente sometidos a un seguimiento temporal, implicando una serie de visitas hospitalarias en las que, por lo general, se recogen una serie de características clínicas, que suponen una recolección de datos que siguen una determinada dinámica longitudinal. Muchas de estas características pueden tener una evolución asociada con una mejoría o empeoramiento de la patología, y conociendo cómo se asocia la evolución con un determinado objetivo, es posible modular o adaptar de manera más eficaz la gestión sanitaria del paciente.

Este comportamiento longitudinal de ciertas variables recogidas en las visitas puede estar asociada con el tiempo hasta la aparición de un determinado evento (fallecimiento, recaída, mejoría, etc.). Los métodos estadísticos clásicos para el análisis de estos datos consistían en procesos de dos etapas, ya fuesen etapas independientes ^[2,3] o etapas relacionadas en las que el resultado de la modelización longitudinal se incluía en el modelo de tiempo hasta el evento ^[4]. Sin embargo, en diversos estudios se han evidenciado una serie de limitaciones y sesgos potenciales al trabajar con estas metodologías ^[5,6,7,8].

Si se aplican análisis por separado, varios estudios mediante técnicas de simulación han identificado evidentes sesgos en los que se incurre al no contemplar que, si los resultados longitudinales y el tiempo hasta el evento se encuentran correlacionados, modelar los dos procesos de resultados de forma independiente puede conducir a estimaciones de tamaño de efecto sesgadas, resultando estos modelos inadecuados para medir la asociación entre la respuesta longitudinal y el tiempo hasta el evento ^[5,6]. La estrategia de dos etapas también incurre en una serie de sesgos que afectan a la eficiencia del modelo y robustez de las estimaciones, infraestimando la verdadera asociación entre ambas respuestas como

consecuencia de un sesgo de dilución, así como una notable sensibilidad a la censura informativa ^[7,8].

Dentro del marco del análisis de supervivencia, se han desarrollado extensiones del modelo de regresión de Cox ^[9] que permiten la modelización de variables tiempo-dependientes que no cumplen la condición de riesgos proporcionales ^[10]. De la misma manera que con los enfoques mencionados previamente, esta metodología infraestima la verdadera asociación entre la respuesta longitudinal y el riesgo de evento, como consecuencia del error experimental derivado de asumir que la tendencia longitudinal se mantiene constante entre las visitas, así como al no contemplar la posible interdependencia entre ambos procesos ^[8].

Dada la notable necesidad de implementar métodos robustos para estimar la asociación entre el comportamiento longitudinal de variables clínicas sobre la ocurrencia de un determinado evento, se han propuesto los denominados modelos conjuntos de verosimilitud o modelos *joint* ^[11,12], que permiten integrar de forma simultánea la información aportada por el componente longitudinal y la información del tiempo hasta el evento. De esta manera, mediante una estructura compartida de efectos aleatorios, se logra estimar la relación entre la tendencia longitudinal de la variable de interés y la ocurrencia del evento a analizar, determinándose así el efecto de dicha dinámica sobre el tiempo hasta que ocurra el evento ^[7]. Mediante este enfoque se incrementa la eficacia y robustez de las estimaciones, reduciendo notablemente los sesgos potenciales de la modelización. Además, en los últimos años se han desarrollado extensiones del modelo *joint* original o clásico, permitiendo estimar modelos conjuntos multivariantes con más de una variable respuesta longitudinal, así como abordar análisis de riesgos competitivos, incluso desde un enfoque multiestado ^[13,14]

Atendiendo a todas las ventajas que ofrece esta metodología estadística, se ha considerado la más adecuada para abordar el problema clínico planteado, en el que se tienen una serie de medidas repetidas de dos biomarcadores de interés, TRECs y KRECs, y se pretende determinar el valor pronóstico del comportamiento longitudinal de ambos sobre el tiempo hasta recaída y/o fallecimiento a los 2 años del trasplante alogénico en los pacientes estudiados.

Desde la perspectiva clínica, existen determinadas patologías que tienen una incidencia elevada en la población y se asocian con un peor pronóstico por lo general y, por lo tanto,

identificar marcadores robustos con un valor pronóstico definido es una necesidad fundamental. Dentro de este grupo de patologías, las enfermedades hematológicas, como leucemias o síndromes mieloproliferativos, son una de las principales causas de morbi-mortalidad en el mundo ^[15]. Para el tratamiento de este tipo de patologías, una de las estrategias terapéuticas que más ha progresado en las últimas décadas es el trasplante alogénico de células madre hematopoyéticas (HSCT) ^[16], más comúnmente conocido como trasplante de médula ósea (*Bone Marrow Transplantation, BMT*).

El término alogénico indica que las células madre hematopoyéticas (HSC) a trasplantar proceden de un donante que puede estar relacionado o no relacionado con el paciente, diferenciándose del trasplante autólogo en el cuál las HSC proceden de la médula ósea del propio paciente. El proceso completo consiste en una primera etapa en la que se somete al paciente a una terapia de acondicionamiento mediante quimioterapia para eliminar las células enfermas, así como liberar espacio para las células sanas trasplantadas, seguida de una segunda etapa consistente en el trasplante propiamente dicho ^[17].

Tras el trasplante, los pacientes, en su mayoría, se encuentran en un estado de inmunodeficiencia crítica, poniéndose de manifiesto que una rápida y eficaz reconstitución de los procesos inmunoproliferativos (en timo y médula ósea) encargados de la respuesta inmune adaptativa se asociará con una mejor recuperación ^[18].

Dicha respuesta inmune está íntimamente ligada a la linfogénesis tanto de linfocitos T como B. Hace tiempo se definieron los denominados TRECs (*T-cell receptor excision circles*) ^[19] que son, a grandes rasgos, pequeños fragmentos de ADN circular que se generan durante el reordenamiento del TCR (receptor de célula T) en el timo durante la producción de linfocitos T, y tienen la característica fundamental de no dividirse durante los procesos mitóticos y, como consecuencia, ser elementos muy estables y conservados. Por ello, la cuantificación de dichos TRECs en sangre periférica mediante técnicas de diagnóstico molecular, como la reacción en cadena de la polimerasa (PCR), permite valorar la eficiencia de la función tímica de una forma bastante precisa, siendo un biomarcador válido para medir la reconstitución de respuesta mediada por linfocitos T ^[20,21].

Para determinar un biomarcador que permita estimar la reconstitución de la respuesta humoral mediada por linfocitos B existe una mayor problemática ^[22]. Siguiendo la misma

línea que con los TRECs, se propusieron los denominados KRECs (*K-deleting recombination excision circles*), que son exosomas o pequeños círculos de ADN procedentes de la recombinación de la cadena ligera de la IgG (receptor de linfocito B) durante la génesis de estos linfocitos en la médula ósea ^[23], y que tienen la misma característica clave de los TRECs de no dividirse durante los procesos mitóticos.

Si bien los TRECs se han evidenciado como biomarcadores robustos de la de la función tímica ^[20] así como estudiado su valor pronóstico tras trasplante alogénico HSCT ^[24,25,26] no ocurre lo mismo con los KRECs, que carecen de evidencia sólida sobre su valor pronóstico tras un trasplante alogénico, sobre todo al analizarlos simultáneamente con los TRECs ^[27]. No obstante, en los últimos años, gracias a los avances en los métodos de cuantificación, comienza a asentarse una base sólida para definir su función como factor pronóstico de la reconstitución inmune adaptativa humoral tras una situación de linfopenia B ^[28].

Por ello, el presente trabajo explora, mediante la aplicación de una metodología estadística novedosa como son los modelos *joint*, el posible valor pronóstico de los TRECs y KRECs sobre distintos desenlaces a los 2 años del trasplante alogénico, en pacientes con diversas hemopatías sometidos a dicho trasplante.

2. HIPÓTESIS Y OBJETIVOS

HIPÓTESIS:

El empleo de los modelos *joint* permitirá evaluar el valor pronóstico de los TRECs y los KRECs en resultados de supervivencia libre de progresión, recaída y muerte sin recaída en pacientes que reciben un trasplante alogénico de médula ósea.

OBJETIVOS:

- El objetivo principal: determinar la posible asociación existente entre el comportamiento longitudinal de los biomarcadores endógenos TRECs y KRECs, tanto de forma univariante como multivariante, y el tiempo libre de progresión (PFS) a los 2 años post trasplante en pacientes sometidos a trasplante alogénico de células madre hematopoyéticas.
- El objetivo secundario: analizar la asociación entre dicha tendencia longitudinal y el riesgo sobre las dos causas específicas que componen el desenlace clínico principal PFS, siendo éstas recaída y muerte sin recaída, a los 2 años del trasplante alogénico.

3. METODOLOGÍA

3.1. DISEÑO DEL ESTUDIO Y MATERIAL

3.1.1 Tipo de estudio, población y base de datos

Estudio de una cohorte observacional, prospectivo, longitudinal y dinámico, en el que se analizan 208 pacientes con diversas enfermedades hematológicas, derivados al servicio de hematología del Hospital Universitario Puerta de Hierro en Madrid entre los años 2009 y 2017. Los pacientes son sometidos a trasplante alogénico de células madre hematopoyéticas (HSCT) como estrategia terapéutica para el tratamiento de la hematopatía. Tras el trasplante fueron sometidos a seguimiento clínico para evaluar periódicamente el estado de salud. El muestreo llevado a cabo, por lo tanto, es no probabilístico, incluyendo la serie consecutiva de casos que recibieron trasplante en ese período de tiempo.

En el trabajo se centra el interés en los biomarcadores TRECs y KRECs, por lo que el objetivo es definir y obtener modelos estimativos y no predictivos, siendo incluidas el resto de las variables contempladas en el estudio como covariables de ajuste. Los niveles de estos dos biomarcadores de interés se obtuvieron mediante cuantificación absoluta por PCR en tiempo real (qPCR) en la plataforma LightCycler 480 de Roche Diagnostics en muestras seriadas recogidas de forma prospectiva en cuatro períodos tras el trasplante, siendo estos: al mes (1m post-HSCT), a los 3 meses (3m post-HSCT), a los 6 meses (6m post-HSCT)), y a los 12 meses (12m post-HSCT). Se aporta también una quinta medida, recogida previamente al HSCT que, siguiendo el criterio clínico, es mantenida en los análisis como una variable de ajuste fija a expensas de la significación estadística que se estime.

Como se evidencia en el párrafo anterior, las medidas longitudinales que se tienen de ambos biomarcadores acotan un período de un año tras el trasplante. Inicialmente se definió un período de seguimiento de un año, con el fin de poder extrapolar de manera más adecuada las conclusiones obtenidas sobre la asociación entre la dinámica de estos biomarcadores y el tiempo hasta el evento. En base a criterio clínico, se decidió ampliar la extensión del período de seguimiento de los 12 hasta los 24 meses, permitiendo incluir en el análisis un mayor número de eventos que si se hubiese limitado a los 12 meses.

Consultando con médicos hematólogos, se asume que el nivel de los biomarcadores pasado el primer año post trasplante, se mantiene estable.

Para ingresar en el estudio, el principal criterio de inclusión establecido es el haber sido sometido a trasplante alogénico HSCT, ya que define no solo la entrada al estudio, si no la condición clave a estudiar. Como criterio de exclusión se estableció excluir aquellos pacientes que no tuviesen, en ningún período post trasplante, medidas de los niveles de TRECs y KRECs, ya que no podrían contribuir con ninguna información al presente análisis. En la Figura 1 se esquematiza, mediante un diagrama de flujo, el número de pacientes incluidos en la base de datos original y el número de pacientes incluidos para el análisis tras la aplicación del criterio de exclusión mencionado. Finalmente se trabaja con un tamaño muestral de 204 pacientes.

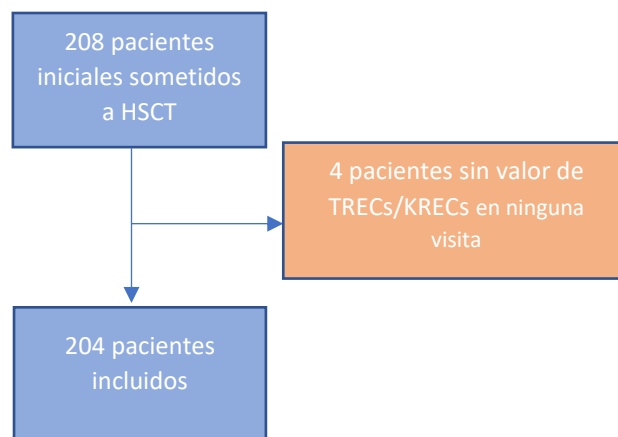


Figura 1. Diagrama de flujo representando la aplicación de los criterios de inclusión y exclusión. 204 pacientes para el análisis final.

La base de datos original incluía un total de 160 variables, incluyendo variables tanto fijas como variables recogidas de forma longitudinal, siendo éstas últimas sobre todo variables clínicas como niveles de elementos asociados a la respuesta inmune (niveles leucocíticos, plaquetarios, de hemoglobina, etc.) que, por criterio clínico, se han excluido del análisis como variables de ajuste. Siguiendo este mismo criterio clínico, la depuración de la base de datos condujo a la exclusión del grueso de variables presentes en la base original, manteniendo aquellas covariables más importantes desde el punto de vista clínico para ajustar los modelos, así como aquellas variables que recogen los tiempos de trasplante, recaída, fallecimiento, etc. Finalmente, las variables que se mantuvieron en la base de datos fueron (Tabla 1):

A)		B)	
TRECS	Variable cuantitativa continua. 4 medidas por paciente	SEXO	Variable categórica dicotómica 0→ Hombre 1 → Mujer
KRECS	Variable cuantitativa continua. 4 medidas por paciente	EDAD AL TRASPLANTE	Variable cuantitativa discreta
C)		PATOLOGÍA HEMATOLÓGICA	Variable categórica politómica 1→ LA/SMD 2→ Aplasia 3→ MM 4→Otros 5→SLP
FECHA DE TRASPLANTE	Variable de fecha	TIPO DE DONANTE	Variable categórica politómica 1→ DNE 2→Haplo 3→ HLAid 4→ TSCU
FECHA DE RECAÍDA	Variable de fecha		
FECHA DE EXITUS	Variable de fecha	ENFERMEDAD INJERTO CONTRA HUÉSPED (EICH)	Variable categórica dicotómica 0→ No 1 → Sí
FECHA FINAL DE SEGUIMIENTO	Variable de fecha		
TIEMPO HASTA RECAÍDA	Variable continua	ACONDICIONAMIENTO PREVIO	Variable categórica dicotómica 0→ Mieloablatoivo 1 → Semi-mieloablatoivo
TIEMPO HASTA EXITUS	Variable continua	TRECS PRE-TRASPLANTE	Variable continua
TIEMPO IDENTIFICANDO CADA MEDIDA	Variable categórica ordinal 1→1m post-TPH 2→3m post-TPH 3→ 6m post-TPH 4→ 24m post-TPH	KRECS PRE-TRASPLANTE	Variable continua

Tabla 1. Variables incluidas en la base de datos una vez depurada. A) Variables longitudinales. B) Variables fijas de ajuste. TPH: Trasplante // LA / SMD : Linfoma agudo / síndrome mielodisplásico // MM: Mieloma múltiple // SLP: Síndrome linfoproliferativo C) Variables de tiempo.

3.1.2 Desenlaces clínicos principales y secundarios

- El desenlace clínico principal es el tiempo libre de progresión (*Progression-Free Survival*, PFS) a los 2 años del HSCT, que se define como el tiempo desde el trasplante hasta la recaída del paciente o la muerte, recogándose el primer evento que ocurra de los dos posibles. Aquellos pacientes que no presenten ninguno de los dos eventos a los 2 años del HSCT, serán considerados censuras no informativas. Utilizando la información aportada por las variables de tiempo se

definió la variable PFS en las bases en formato ancho para los modelos de supervivencia, recogiendo en una variable dicotómica el evento y la censura cada paciente (0 → Censura; 1 → Evento).

- El desenlace clínico secundario deriva de desgarnar el principal en las dos causas específicas que lo definen, recaída y fallecimiento sin recaída (*Non-Relapse Mortality* NRM) a los 2 años del HSCT. Se define como evento principal la recaída, siendo el fallecimiento sin recaída el evento que compite con dicho evento principal. De la misma manera que para el desenlace principal, se definen las variables dicotómicas de censura para cada evento específico.

3.2. METODOLOGÍA ESTADÍSTICA

3.2.1 Análisis descriptivo. Imputaciones y transformación de TRECs y KRECs.

Al trabajar con datos clínicos longitudinales, recogidos a lo largo de una serie de visitas, suele ser habitual no tener completas todas las observaciones, teniendo valores faltantes o *missing* en algunas de ellas. Esto conlleva no sólo pérdida de información, si no en el caso del análisis de supervivencia, la pérdida de eventos al no tener información de la covariable de interés en el momento de ocurrencia del evento. Para corregir la ausencia de valores en el periodo de fallecimiento, se ha llevado a cabo una imputación LOCF (*Last Observation Carried Forward*) arrastrando el último valor recogido de los biomarcadores para aquellos pacientes que fallecían al año del trasplante.

Uno de los requisitos para poder aplicar la metodología de los modelos *joint* es que la base de datos en formato largo para los modelos longitudinales y la base de datos en formato ancho usada para el modelo de Cox han de tener el mismo tamaño muestral, siendo además con los mismos pacientes, ya que la intensidad de la asociación se estima en base a los efectos aleatorios compartidos, que incluyen la identificación de cada paciente (*intercept* aleatorio). Para corregir la ausencia de valores en las medidas pre trasplante de TRECs y KRECs, se ha llevado a cabo una imputación múltiple por el valor más probable, en base a covariables seleccionadas bajo criterio clínico. Los resultados de las imputaciones de datos aplicadas en el trabajo se han verificado con el equipo clínico compuesto por dos hematólogos.

Para resumir la información aportada por las variables recogidas en el estudio, se reporta:

- Para covariables categóricas, las frecuencias absolutas y frecuencias relativas de cada categoría definida en la variable.
- Para covariables cuantitativas fijas, un sumario de estadísticas incluyendo la media y desviación estándar, primer, segundo (mediana) y tercer cuartil, y el valor mínimo y máximo.
- Para las variables longitudinales TRECs y KRECs, en cada medida pre y post TPH (trasplante) la mediana y el primer y tercer cuartil, acotando así los límites en los que se encuentran el 50% de los valores observados. Se reporta tanto para la variable cruda como transformada. Se incluye también un gráfico de densidad estratificado por las medidas pre y post TPH para los valores de las variables longitudinales crudas, así como transformadas.
- Para las variables de evento o censura, se reportará en cada medida post TPH las frecuencias absolutas de incidencia, así como la frecuencia relativa sobre el total de los pacientes en riesgo en ese punto.

Para describir la relación entre el comportamiento longitudinal de los biomarcadores y la ocurrencia de los eventos, se elaboraron gráficos de cajas estratificados por las variables PFS, recaída y NRM, tanto para TRECs como para KRECs, ambos tras la transformación logarítmica. Se reporta también la representación gráfica de la curva de supervivencia de Kaplan-Meier a los 2 años del trasplante para el desenlace principal PFS, así como un gráfico de perfil para representar la tendencia individual de TRECs y KRECs post TPH en 5 pacientes seleccionados al azar.

3.2.2 Modelos Mixtos para el análisis de datos longitudinales

Los modelos mixtos son una generalización del modelo lineal general, que permite contemplar la correlación existente entre las observaciones. A diferencia de los modelos lineales generales, donde únicamente se contemplan factores fijos y covariables numéricas, y se asume que el valor esperado de la variable respuesta se puede expresar como una combinación lineal de las covariables explicativas, los modelos lineales mixtos permiten incluir un factor aleatorio que identifica cada nivel o unidad experimental en la que se repite en más de una ocasión la variable respuesta ^[29].

Las medidas repetidas o datos longitudinales presentan unas determinadas características que requieren de un tratamiento estadístico específico. Como ya se ha mencionado, el aspecto fundamental es la correlación entre ellas, que implica la adición de un componente de variabilidad dentro de las medidas del mismo sujeto, que no se contempla al asumirse independencia entre observaciones.

Los modelos lineales tradicionales requieren de ciertas asunciones o supuestos indispensables, que básicamente implican la independencia de los términos de error, que han de seguir una distribución normal centrada en 0 y con una varianza σ^2 constante (homocedasticidad). Este enfoque permite evaluar una relación lineal entre variable respuesta y un conjunto de variables explicativas. La expresión de estos modelos es la siguiente:

$$y_i = X_i\beta + \varepsilon_i; \quad i = 1, \dots, n$$
$$\varepsilon_i \sim N(0, \sigma^2); \quad y_i \sim N(X_i\beta, \sigma^2)$$

Donde y_i es la variable respuesta en el individuo i -ésimo, β el vector de coeficientes asociados a X_i , que es el vector de covariables fijas, y ε_i el término de error, con distribución normal de media 0 y varianza σ^2 constante.

Al añadir un componente longitudinal, como son medidas recogidas a lo largo de una secuencia temporal, se quebranta las asunciones de independencia y homocedasticidad de los residuos, por lo que no se puede aplicar esta metodología.

Para integrar esta condición en un modelo estadístico de regresión, se incluyen uno o varios factores o efectos aleatorios en el modelo. Esta inclusión de efectos fijos junto con efectos aleatorios es lo que otorga el nombre de modelos mixtos a estos modelos de regresión. Los factores fijos serán las variables incluidas en el modelo que sean incluidas a propósito y que incluyan todos los niveles sobre los que se quiere analizar, que explican la variabilidad de la variable respuesta. En el caso del presente estudio, un ejemplo de efecto fijo sería el sexo de los pacientes, o la edad al trasplante.

Los efectos aleatorios, por otra parte, organizan la variabilidad existente, representando una proporción de todos los niveles posibles de una determinada variable, y recogiendo las posibles fuentes de heterogeneidad no controladas por el experimentador. En el caso de este trabajo, hay dos efectos aleatorios fundamentales, uno referente al término

independiente aleatorio, que identifica a cada individuo y sus medidas longitudinales, y otro referente a la pendiente aleatoria, que permite indicar que cada sujeto tiene una tendencia longitudinal de evolución diferente del resto de sujetos.

El modelo mixto incluyendo término independiente aleatorio y pendiente aleatoria se especificaría de forma matricial con la siguiente notación:

$$y_i(t) = X_i(t)\beta + Z_i(t)b_i + \varepsilon_i(t); i = 1, \dots, n$$

$$b_i \sim N(0, G); \varepsilon_i \sim N(0, \sigma^2 I_{n_i})$$

Donde $X_i(t)$ es la matriz diseño para los efectos fijos, β es el vector de coeficientes asociados a los efectos fijos, $Z_i(t)$ la matriz diseño para los efectos aleatorios, y b_i el vector de efectos aleatorios asociados a cada nivel. Este vector de efectos aleatorios sigue una distribución normal cuya variabilidad viene dada por la matriz de varianzas-covarianzas G , que recoge la variabilidad asociada a los efectos aleatorios incluidos y su covarianza. El vector de los términos de error ε_i ya no presenta homocedasticidad, siendo el componente de variabilidad el producto de la matriz identidad I_{n_i} y varianza residual σ^2 .

3.2.3 Análisis de supervivencia para tiempo hasta evento.

El análisis de supervivencia son un conjunto de técnicas estadísticas aplicadas cuando se modeliza el tiempo desde un punto inicial (momento en el que el paciente comienza a estar en riesgo de tener el evento) hasta un tiempo en el que se observa el evento (fallo), conocido como tiempo de supervivencia o fallo. El seguimiento de un paciente a lo largo de un determinado período da margen para la aparición del evento, dependiendo siempre del tipo de evento, así como de la extensión de ese período de seguimiento. De esta forma, se puede estimar si el tiempo hasta la aparición de dicho evento está relacionado con otras variables explicativas, estimando cómo modifican estas variables el riesgo de que suceda el evento estudiado.

Este conjunto de técnicas estadísticas ha ganado popularidad en las últimas décadas por varios aspectos, pero, sobre todo, por las ventajas que ofrece frente a otras técnicas inferenciales para el tratamiento de lo que se denominan datos censurados ^[10]. Los pacientes incluidos en un estudio pueden tener tiempos de entrada al estudio diferentes, así como períodos de seguimiento diferentes, ya sea por abandono o por otras razones, lo

que condiciona la aparición o no del evento en dichos pacientes. Estos casos son los que se conocen como censura, cuando un paciente abandona (o finaliza) el estudio sin haber presentado el evento (censura por la derecha), si el paciente ingresa en el estudio habiendo presentado ya el evento (censura por la izquierda), o censura por intervalo cuando no se conoce el punto temporal exacto en el que se censura al sujeto ^[10]. El tratamiento de los datos con el enfoque del análisis de supervivencia permite contabilizar el tiempo que se mantuvieron en el estudio estos pacientes, ya que se sabe que sobrevivieron un tiempo T y que no presentaron el evento, pudiendo presentarlo en un futuro, lo que supone una estimación más realista y ajustada de la función de riesgo que si se estimase por alguna otra metodología estadística ^[30]. Es importante remarcar que esta censura ha de ser *no informativa*, es decir, han de ser independientes las variables aleatorias asociadas al tiempo de fallo y las variables de censura ^[31].

La función de supervivencia, $S(t)$, es la probabilidad ($0 \leq S(t) \leq 1$), de que el tiempo hasta el evento (T) de un sujeto sea mayor que un tiempo t . Su notación es:

$$S(t) = P(T > t)$$

Esta función de supervivencia se define también como $1 - F(t)$, siendo $F(t)$ la función de distribución, que indica la probabilidad de que el tiempo de supervivencia sea menor o exactamente igual a t .

$$F(t) = P(T \leq t)$$

Al tratarse generalmente la variable tiempo hasta el evento de una variable continua, se especifica de la siguiente manera su función de densidad:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

Que determina la probabilidad de que el tiempo de supervivencia ocurra en torno a un instante t .

Esta función de supervivencia $S(t)$ indica la probabilidad de no presentar el evento antes de un tiempo determinado, pero, en el ámbito biomédico, suele ser también relevante estimar el riesgo de presentarlo en un momento definido, para así poder determinar qué variables modulan ese riesgo, determinando si se tratan de factores de riesgo o factores de protección frente al evento. Este riesgo se puede determinar mediante la función de

riesgo o *hazard* instantáneo $h(t)$, que determina el riesgo de presentar el evento en un tiempo t específico, con la condición de no haberlo presentado en el punto $t-1$. Esta función de *hazard* instantáneo es una tasa y se define como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

y se relaciona con la función de supervivencia y de densidad mediante la siguiente expresión:

$$h(t) = \frac{f(t)}{S(t)}$$

Para modelizar la posible relación entre una o varias covariables y la tasa de riesgo asociada al evento, se puede utilizar el modelo de riesgos proporcionales de Cox ^[2], que es un modelo de tipo semi-parámétrico, que únicamente requiere la condición de que la relación entre los riesgos de los distintos grupos se mantenga constante en el tiempo ^[10].

La expresión matemática del modelo de Cox PH es:

$$h_i(t/Z) = h_0(t) \exp(\beta_1 Z_{i1} + \beta_2 Z_{i2} + \dots + \beta_k Z_{ik}); \quad i = 1, \dots, n$$

Donde Z es el vector de covariables (Z_1, Z_2, \dots, Z_k) , y β es el vector de parámetros de regresión $(\beta_1, \beta_2, \dots, \beta_k)$. $h_0(t)$ es el riesgo basal, que refiere el riesgo cuando todas las covariables tienen valor nulo.

3.2.4 Análisis de supervivencia con eventos competitivos.

El término riesgos competitivos hace referencia a determinadas situaciones en las que los sujetos analizados pueden presentar más de un evento, pero uno de ellos es el evento de interés y el otro evento, si tiene lugar, impide la ocurrencia del evento establecido como principal ^[9]. Existen varias aproximaciones para el tratamiento de este tipo de datos, destacando el modelo de Cox para cada riesgo por causa específica ^[32], el modelo de riesgos aditivos de Aalen ^[33], o el modelo de Fine y Gray basado en las funciones de incidencia acumulada ^[34].

Existe un cuarto enfoque, que ha sido el aplicado en este trabajo, que consiste en modelizar a partir de modelos multiestado sin eventos intermedios. En el marco de este tipo de modelos, se parte de un estado inicial, existiendo posibles transiciones desde dicho

estado inicial a X estados “absorbentes”, como por ejemplo la recaída de paciente o el fallecimiento ^[35], siendo estos estados finales ya que no se contemplan transiciones intermedias. Introduciendo una breve notación matemática, el modelo multiestado se especifica de la siguiente manera:

$$h_{jk}(t) = y_{ijk}(t) \exp(\beta_{jk} Z_i) h_{0jk}(t);$$
$$j, k = 1, 2, \dots, m; j \neq k; i = 1, 2, \dots, n$$

Refiriendo j y k los diferentes estados posibles e i el individuo. β_{jk} es el vector de coeficientes para la transición del estado j al estado k , Z_i el vector de covariables para el individuo i y $h_{0jk}(t)$ el riesgo basal en la transición de j a k . Se introduce un nuevo componente, que es $y_{ijk}(t)$, que valdrá 1 para el individuo i -ésimo si en el tiempo t está en el estado j y en riesgo de experimentar transición al estado k . Este enfoque, por lo tanto, permite estimar el riesgo para cada transición específica entre el estado original y los estados “absorbentes”.

3.2.5 Modelos conjuntos de verosimilitud (modelos *joint*)

En el análisis de datos en biomedicina, las variables recogidas en el fichero de datos se pueden clasificar atendiendo a su naturaleza como:

- Variables exógenas: Aquellas covariables externas al sujeto, como factores económicos, ambientales o sociales. Para este tipo de variables, el uso de las estrategias derivadas de la extensión del modelo de Cox resulta en estimaciones con un riesgo de sesgo menos importante.
- Variables endógenas: Son aquellas covariables cuyos niveles están determinados por el organismo del propio sujeto. Este grupo abarca los denominados biomarcadores, tales como nivel de glucosa, colesterol, niveles enzimáticos, y, en este estudio, los niveles de TRECs y KRECs circulantes. Este tipo de variables son, en muchos casos, dependientes del proceso de supervivencia, ya que variaciones en sus niveles conducirán a un mayor o menor riesgo de experimentar el evento estudiado. Por ello, la aplicación de modelos de Cox para variables tiempo-dependientes, así como análisis en dos etapas, implica una serie de sesgos tanto en la estimación como en la metodología, que conduce a resultados con una validez limitada.

La principal limitación que presentan los métodos clásicos surge a la hora de trabajar con variables tiempo-dependiente endógenas, ya que no permiten explicar la endogeneidad de éstas ^[7], así como no contemplan la posible interdependencia entre el proceso longitudinal y el de supervivencia ^[8].

Como estrategia para modelizar de forma más ajustada y acertada la asociación de variables endógenas tiempo-dependientes sobre el riesgo de fallo para un determinado evento, se han desarrollado los modelos *joint* ^[7,11,12]. La principal ventaja que aportan estos modelos es la capacidad de poder estimar de forma simultánea los parámetros que describen el proceso longitudinal para la variable tiempo-dependiente y los que definen el proceso de supervivencia. De esta manera, se obtienen estimaciones más precisas de la intensidad de la relación entre el biomarcador y el riesgo de fallo. Se describe tanto la evolución del biomarcador en el tiempo mediante un modelo longitudinal de efectos mixtos, como la asociación de dicha evolución sobre el estado del sujeto. El marco de los modelos *joint* requiere:

- Un submodelo de efectos mixtos que modele la respuesta longitudinal del biomarcador:

$$y_i(t) = m_i(t) + \varepsilon_i(t)$$

$$\varepsilon_i(t) \sim N(0, \sigma^2)$$

El componente $m_i(t)$ recoge el valor real y no observado del biomarcador en el tiempo t , englobando tanto la parte de efectos fijos ($X_i(t)\beta$) como la parte de efectos aleatorios ($Z_i(t)b_i$)

$$m_i(t) = X_i(t)\beta + Z_i(t)b_i$$

- Un segundo submodelo de Cox de riesgos proporcionales, en el que no se incluye la variable tiempo-dependiente, si no que se incluyen las variables de ajuste y se modeliza el tiempo hasta el evento para cada paciente:

$$h_i(t / M_i(t)) = h_0(t) + \exp(y^T Z_i + \alpha m_i(t))$$

donde el riesgo en el tiempo t está condicionado al historial longitudinal de la variable tiempo-dependiente $M_i(t) = \{m_i(s), 0 \leq s < t\}$. Asociado a este componente, se encuentra el denominado coeficiente de asociación (α) que

cuantifica la intensidad de la asociación entre el biomarcador longitudinal y el riesgo sobre el fallo. Z_i hace referencia al vector de covariables en basal.

- Finalmente, se modelizan de forma simultánea ambos submodelos, estimando la asociación entre la respuesta longitudinal y el riesgo sobre el tiempo de fallo. La especificación del modelo sería la siguiente, propuesta por Tsiatis & Davidian en 2004 y conocido bajo el nombre de modelos de efectos aleatorios compartidos (*shared-random effects*), siendo la metodología más extendida de los modelos *joint* [6]:

$$p(y_i, T_i, \delta_i) = \int p(y_i | b_i) \{h(T_i | b_i)^{\delta_i} S(T_i | b_i)\} p(b_i) db_i;$$

$$i = 1, 2, \dots, n$$

Donde $p(y_i, T_i, \delta_i)$ refiere la función de densidad de la respuesta longitudinal (y_i), el tiempo observado hasta el evento para el paciente i -ésimo (T_i) y la variable que recoge el evento y censura (δ_i). En todos los componentes de la ecuación se puede apreciar que se incluye como condicionante el vector de efectos aleatorios b_i que explica todas las interdependencias, es decir, la relación o dependencia entre la respuesta longitudinal y el riesgo de evento se describe por una estructura compartida de parámetros de efectos aleatorios sujeto-específicos, asumiéndose total independencia en el resto de las respuestas. Dicho vector de efectos aleatorios permite introducir la historia longitudinal como una covariable en el modelo de riesgos proporcionales, definiendo una estructura de asociación que determinará la relación entre las medidas repetidas y el riesgo de evento [10].

Para la modelización y obtención de las estimaciones, se han desarrollado estrategias y funciones en softwares estadísticos para estimar los modelos *joint* desde el enfoque frecuentista, mediante la maximización de la función conjunta de verosimilitud [7,36], así como mediante aproximación Bayesiana utilizando los métodos de cadenas de Markov-Montecarlo (MCMC) [37]. En este trabajo se ha optado por la aproximación Bayesiana, asumiendo una distribución a priori no informativa, para evitar posibles problemas de convergencia, y también porque permite, entre otros aspectos, la modelización

multivariante con más de una respuesta longitudinal. La aproximación bayesiana se puede especificar como:

$$\pi(\theta, b_i | D_n) \propto p(D_n | b_i, \theta) \pi(\theta, b_i) = p(D_n | b_i, \theta) p_b(b_i | \theta) \pi(\theta);$$

$$\theta = (\theta_y^T, \theta_t^T, \theta_b^T); D_n = \{T_i, \delta_i, y_i; i = 1, \dots, n\}$$

Donde D_n es un vector que recoge la información muestral, θ un vector de parámetros desconocidos que incluye los parámetros longitudinales (θ_y^T), de tiempo hasta el evento (θ_t^T) y los componentes de la matriz de efectos aleatorios (θ_b^T). Por lo tanto, $\pi(\theta, b_i | D_n)$ refiere la distribución a posteriori, $p(D_n | b_i, \theta)$ la función de verosimilitud y $\pi(\theta, b_i)$ la distribución a priori. $p_b(b_i | \theta)$ refiere la distribución a priori de la estructura compartida de efectos aleatorios.

A) Modelos joint para más de una respuesta longitudinal

El gran interés que han despertado los modelos *joint* ha promovido el desarrollo de extensiones de los modelos *joint* estándar para abordar situaciones en las que se quieren realizar inferencias con más de una respuesta longitudinal. Entre estas extensiones, destacan diversas estrategias para estimar la asociación entre varios biomarcadores longitudinales y el riesgo sobre un evento determinado ^[38]. En el caso de este estudio, por ejemplo, se tienen dos biomarcadores (TRECs y KRECs), y el interés reside no sólo en conocer cómo se asocia uno y otro por separado con el desenlace clínico de interés, si no cómo se comporta dicha relación en un modelo multivariante, ajustando ambos biomarcadores.

La estrategia más extendida actualmente consiste en un método de dos etapas o fases ^[39]:

- Primera etapa: Mediante métodos MCMC se ajusta un modelo mixto multivariante con las n variables respuesta longitudinales, obteniéndose una muestra de la distribución conjunta a posteriori con la siguiente notación:

$$\{\theta_y^{(m)}, b^{(m)}; m = 1, \dots, M\} \sim \{\theta_y, b | y_{li}, \dots, y_{ki}\}$$

Siendo $\theta_y^{(m)}$ el conjunto de parámetros incluidos en el modelo multivariante, incluidos los de la distribución de efectos aleatorios (b).

- Segunda etapa: Usando la muestra seleccionada en la primera etapa, se obtiene una muestra de la distribución conjunta a posteriori de los parámetros del

modelo de supervivencia, siendo en este caso $\theta_t^{(m)}$ el conjunto de parámetros incluidos en dicho submodelo de supervivencia:

$$\{\theta_t^{(m)}; m = 1, \dots, M\} \sim \{\theta_t | T_i, \delta_i, b^{(m)}, \theta_y^{(m)}\}$$

Donde T_i es el vector de tiempos observados hasta evento y δ_i el vector de censura, ambos para el individuo i -ésimo.

Sin embargo, esta aproximación tiene sus contras, y es que las estimaciones se encuentran sesgadas como consecuencia de no trabajar con la distribución conjunta completa ^[6]. Para subsanar este sesgo en las estimaciones, se ha desarrollado una corrección basada en una ponderación de las muestras seleccionadas, que implica asumir que las muestras seleccionadas $\{\theta_t^{(m)}, \theta_y^{(m)}, b^{(m)}; m = 1, \dots, M\}$ son muestras ponderadas de la distribución conjunta a posteriori, estimando el peso al que se asocian según la siguiente expresión ^[40]:

$$\varpi^{(m)} = \frac{p(\theta_t^{(m)} | T, \delta, \theta_y^{(m)}, b^{(m)}) p(\theta_y^{(m)}, b^{(m)} | y, T, \delta)}{p(\theta_t^{(m)} | T, \delta, \theta_y^{(m)}, b^{(m)}) p(\theta_y^{(m)}, b^{(m)} | y)}$$

Donde el numerador recoge la expresión de la distribución conjunta a posteriori completa, y el denominador la distribución conjunta a posteriori de las muestras seleccionadas en las dos etapas del proceso, previo a la corrección.

B) Modelos joint para riesgos competitivos con modelos multiestado

Otra de las limitaciones que presentaban los modelos *joint* estándar era la imposibilidad de modelizar situaciones en las que se tenía más de un evento en la estructura de supervivencia, ya fuesen eventos recurrentes o eventos competitivos. Para extender la metodología *joint* y abarcar estos aspectos, se han desarrollado diversas aproximaciones, entre las que se hará énfasis en las técnicas para integrar la información de un modelo de efectos mixtos con la información aportada por un modelo de supervivencia multiestado sin eventos intermedios ^[13,14].

La principal diferencia reside en el ajuste del submodelo multiestado de supervivencia, En la segunda etapa del proceso, es necesario especificar cada transición específica entre los diferentes estados que se contemplen, de tal manera que se estiman los parámetros para cada transición específica, condicionados al comportamiento del biomarcador

longitudinal a través de la estructura de efectos aleatorios compartida b_i . Si se transita de un estado j en un tiempo determinado al estado k , el riesgo de la transición en el tiempo t se especifica según la siguiente expresión ^[13]:

$$h_{jk}^i(t) = h_{jk0(t)} \exp \left\{ Z_{jk,i}^T y_{jk} + W_{jk,i}(b_i, t)^T \alpha_{jk} \right\};$$

$$j, k = 1, 2, \dots, m; j \neq k; i = 1, 2, \dots, n$$

Donde $h_{jk0(t)}$ es el riesgo basal para dicha transición $j \rightarrow k$, y_{jk} el vector de covariables y $Z_{jk,i}^T$ el vector de los parámetros de regresión estimados para cada covariables para esa transición $j \rightarrow k$, $W_{jk,i}(b_i, t)^T$ define la estructura de interdependencia entre el modelo longitudinal y el modelo multiestado a partir de los efectos aleatorios compartidos b_i , y α_{jk} es el componente que cuantifica la asociación entre la dinámica del biomarcador longitudinal y el riesgo de transición.

De esta manera, las estimaciones de los parámetros de asociación entre la respuesta longitudinal y las diferentes transiciones permitirá obtener un resultado para cada transición final, diferenciando la ocurrencia de ambos eventos competitivos, y permitiendo realizar inferencias sobre cada transición final específica.

Software utilizado

Para la realización del análisis descriptivo de las variables, así como para determinados aspectos de la depuración de la base de datos original, se utilizó el programa estadístico Stata en su versión 16.0 (*StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC.*). Para todo el análisis inferencial, incluyendo la programación y aplicación de los modelos *joint*, se utilizó el programa Rstudio en su versión 1.2.1335. (*RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.*).

4 RESULTADOS

4.1 Análisis descriptivo

El estudio incluye 204 pacientes diagnosticados con una hematopatía, sometidos a un trasplante alogénico de células madre hematopoyéticas, de los que se obtienen muestras seriadas de TRECs y KRECs en 4 periodos tras el trasplante, y una medida antes de someterse al trasplante.

El fichero de datos original presentaba una serie de valores faltantes en las medidas repetidas y pre trasplante de TRECs y KRECs distribuidas de la siguiente manera:

- Al año tras trasplante, 40/42 (95.24%) no tenían medida de TRECs, y/o de KRECs en el periodo en el que fallecían.
- 15/204 pacientes (7.35%) carecían de la medida pre trasplante de TRECs y 13/204 (6.37%) les faltaba la medida pre trasplante de KRECs.

Para facilitar la elaboración y aplicación de los modelos *joint*, se decidió con el aval clínico de los hematólogos realizar dos imputaciones independientes:

- Una imputación LOCF (*Last Observation Carried Forward*) para los 40 pacientes que no presentaban nivel de ambos biomarcadores en el periodo de fallecimiento. La imputación consistió en correr el valor de la última medida que se tuviese para los pacientes que fallecían dentro del margen de un año tras el trasplante. Se excluyó de esta imputación a los pacientes que no tuviesen valor del biomarcador en la última medida una vez extendido el período de seguimiento hasta 24 meses, ya que no se podía asegurar que no fuese a sesgar los resultados el asumir que una medida tomada 1,3 o 6 meses tras el trasplante pudiese seguir manteniéndose en niveles similares en un período de más de seis meses de diferencia.
- Imputación múltiple por el valor más probable: Esta imputación se realizó sobre los valores *missing* de 15 medidas pre trasplante de ambos biomarcadores, integrando el valor más probable de dicha medida en base a covariables de interés seleccionadas por los clínicos. Las variables seleccionadas para ajustar fueron el sexo, la edad y la patología hematológica, ya que son las únicas covariables recogidas antes del trasplante, y permiten estimar el valor más plausible de TRECs y KRECs pre trasplante en pacientes con características similares. La imputación

se aplicó sobre los valores crudos o naturales de las variables, y posteriormente se aplicó la transformación logarítmica.

Tras la imputación, se comprobó que los valores de las variables antes y después de aplicar el método de imputación no se habían visto muy modificadas, asegurando que las variaciones eran mínimas ya que el porcentaje de valores imputados fue relativamente bajo.

En la tabla 2 se muestra una descriptiva de las variables categóricas referentes a las características basales de los pacientes (A), así como un sumario de estadísticas de la edad, única variable fija continua a excepción de las referentes a los biomarcadores de interés (B).

A)

VARIABLE		N (%)
SEXO		
FEMENINO	84 (41.17%)	
MASCULINO	120 (58.83%)	
TIPO TPH		
DNE	40 (19.61%)	
HAPLO	10 (4.90%)	
HLAID	94 (46.08%)	
TSCU	60 (29.41%)	

B)

VARIABLE	MIN	MEDIA (SD)	MEDIANA (Q1; Q3)	MAX
EDAD (AÑOS)	16	43.84 (12.91)	45 (33; 55)	68

VARIABLE		N (%)
PATOLOGÍA HEM.	LA/SMD	148 (72.55%)
	APLASIA	1 (0.49%)
	MM	14 (6.86%)
	SLP	32 (15.69%)
	OTROS	9 (4.41%)
EICH	NO	138 (67.65%)
	SÍ	66 (32.35%)
ACONDICIONAMIENTO	MIELOABLATIVO	104 (50.98%)
	SEMI-MIELOABLATIVO	100 (49.02%)

Tabla 2. Análisis descriptivo de las variables fijas. **A)** Variables categóricas. Se indica la frecuencia absoluta y entre paréntesis la frecuencia relativa. **B)** Variable continua.

En la muestra de pacientes analizada, hay un porcentaje ligeramente mayor de hombres que de mujeres, así como una prevalencia mucho mayor de hemopatías leucémicas agudas (LA) y síndromes mielodisplásicos (SMD) frente al resto de hemopatías contempladas. El mayor porcentaje de donantes para el trasplante alogénico HSCT fueron HLA idénticos frente al resto, y el acondicionamiento previo al trasplante se distribuye en partes iguales entre los que fueron sometidos a un acondicionamiento mieloablativo o a un acondicionamiento semi mieloablativo. Tras el trasplante, un 32.35% de los pacientes desarrollaron enfermedad de injerto contra huésped, es decir, 1 de cada 3 pacientes

aproximadamente. En el momento del trasplante, la mediana de edad de los pacientes era de 45 años, encontrándose el 50% de las edades entre los 33 y los 55 años.

Atendiendo al comportamiento longitudinal de los biomarcadores TRECs y KRECs en la muestra del estudio, éstos presentaban una marcada variabilidad, por lo que se decidió trabajar con una transformación logarítmica ($\log(I+X)$) de ambas variables con el objetivo de reducir dicha dispersión. En la Figura 2 se muestra la representación gráfica de la función de densidad de la variable TRECs (incluyendo la medida pre trasplante) en su versión cruda o natural, y en su versión transformada, y la Tabla 3 recoge los valores numéricos de dicha variable expresados en mediana (Q1; Q3) para cada medida. Analizando la dinámica global de los TRECs en los 5 periodos estudiados, se observa una reducción entre los niveles cuantificados antes y a un mes del trasplante, manteniéndose bajos dichos niveles hasta 3 meses tras el trasplante, seguido de un incremento en la reconstitución de la capacidad inmune adaptativa de los linfocitos T en el período de 3 a 24 meses tras el trasplante.

En cuanto al biomarcador asociado a la reconstitución de la respuesta mediada por linfocitos B, se muestra en la Figura 3 las gráficas de densidad asociadas a la variable KRECs en crudo y transformada, junto con la Tabla 4 que recoge el análisis descriptivo en cada observación para esta variable (incluyendo la medida pre trasplante). A diferencia de los TRECs, antes del trasplante los niveles cuantificados de copias de KRECs eran muy bajos, tendencia que se mantenía al mes del trasplante. Sin embargo, entre el mes y los 3 meses tras el trasplante, se daba un incremento notable en la expresión de este biomarcador, estabilizándose dichos niveles tras los 6 meses después del HSCT.

Atendiendo a la información relativa al tiempo hasta el evento, se pueden diferenciar tres desenlaces clínicos, el referente al objetivo principal (tiempo libre de progresión o PFS), y, desgranando dicho desenlace, tiempo hasta recaída y tiempo hasta muerte sin recaída (NRM). En la Figura 4 se muestra la curva de supervivencia de Kaplan-Meier de la serie de casos estudiada para el objetivo principal PFS a los 24 meses.

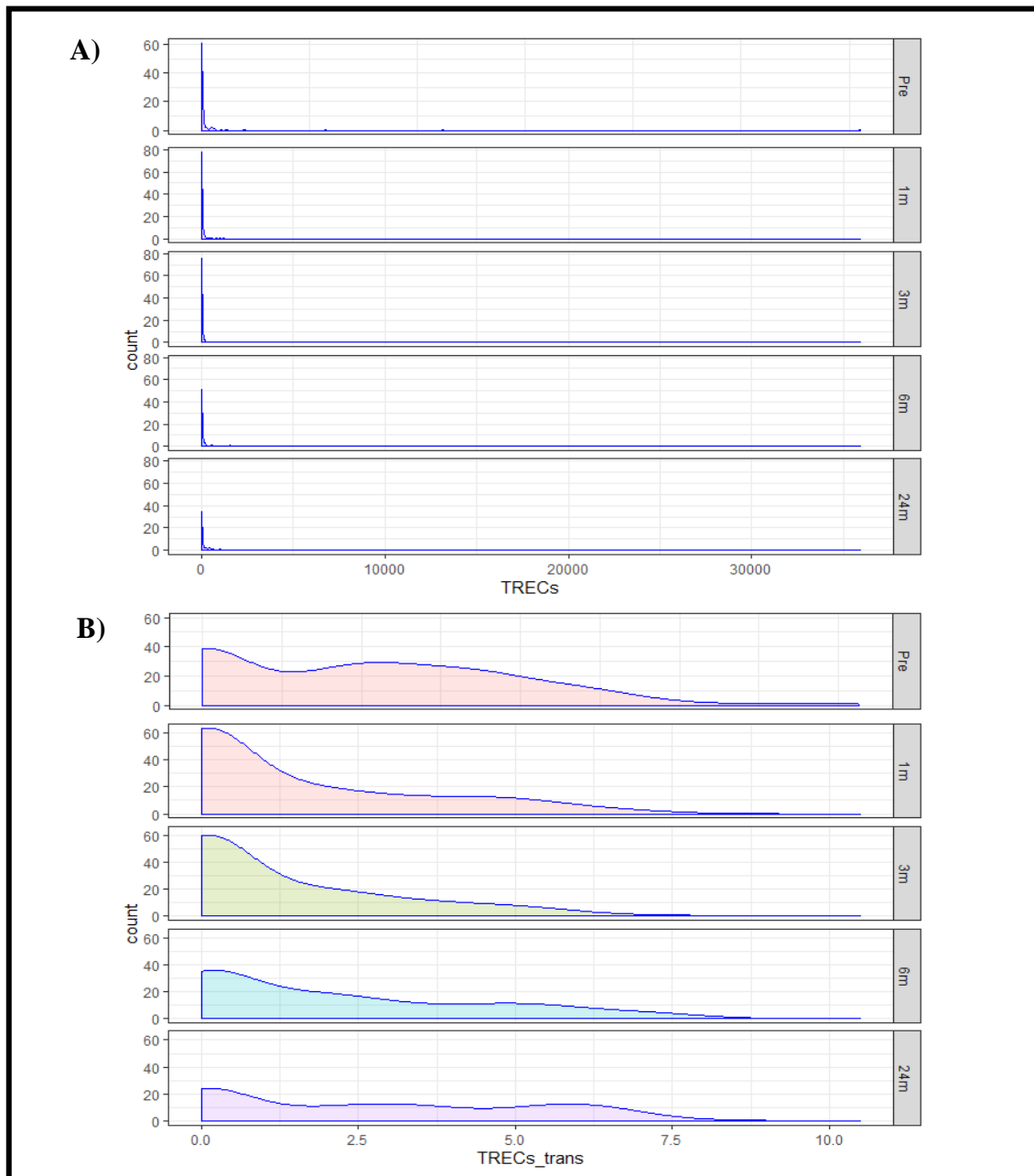


Figura 2. Gráficos de densidad de los valores de la variable TRECs en crudo (A) y tras transformación logarítmica $\log(1+\text{TRECs})$ (B), en los 5 períodos en los que se tomaron medidas.

VARIABLE	PRE-TPH	1M POST TPH	3M POST TPH	6M POST TPH	24M POST TPH
	MEDIANA (Q1; Q3)	MEDIANA (Q1; Q3)	MEDIANA (Q1; Q3)	MEDIANA (Q1; Q3)	MEDIANA (Q1; Q3)
TRECS	10.3 (0; 62.91)	0.01 (0; 8.09)	0 (0; 7.95)	3.95 (0, 39.55)	9.58 (0; 156.75)
LOG (1+TRECS)	2.425 (0; 4.16)	0.01 (0; 2.21)	0 (0; 2.19)	1.60 (0; 3.70)	2.36 (0; 5.06)

Tabla 3. Análisis descriptivo de la tendencia longitudinal de TRECs y $\log(1+\text{TRECs})$ a lo largo de los 5 periodos estudiados, refiriendo la mediana de la distribución en cada periodo, así como los valores de Q1 y Q3 que acotan el 50% de los valores en dicho periodo.

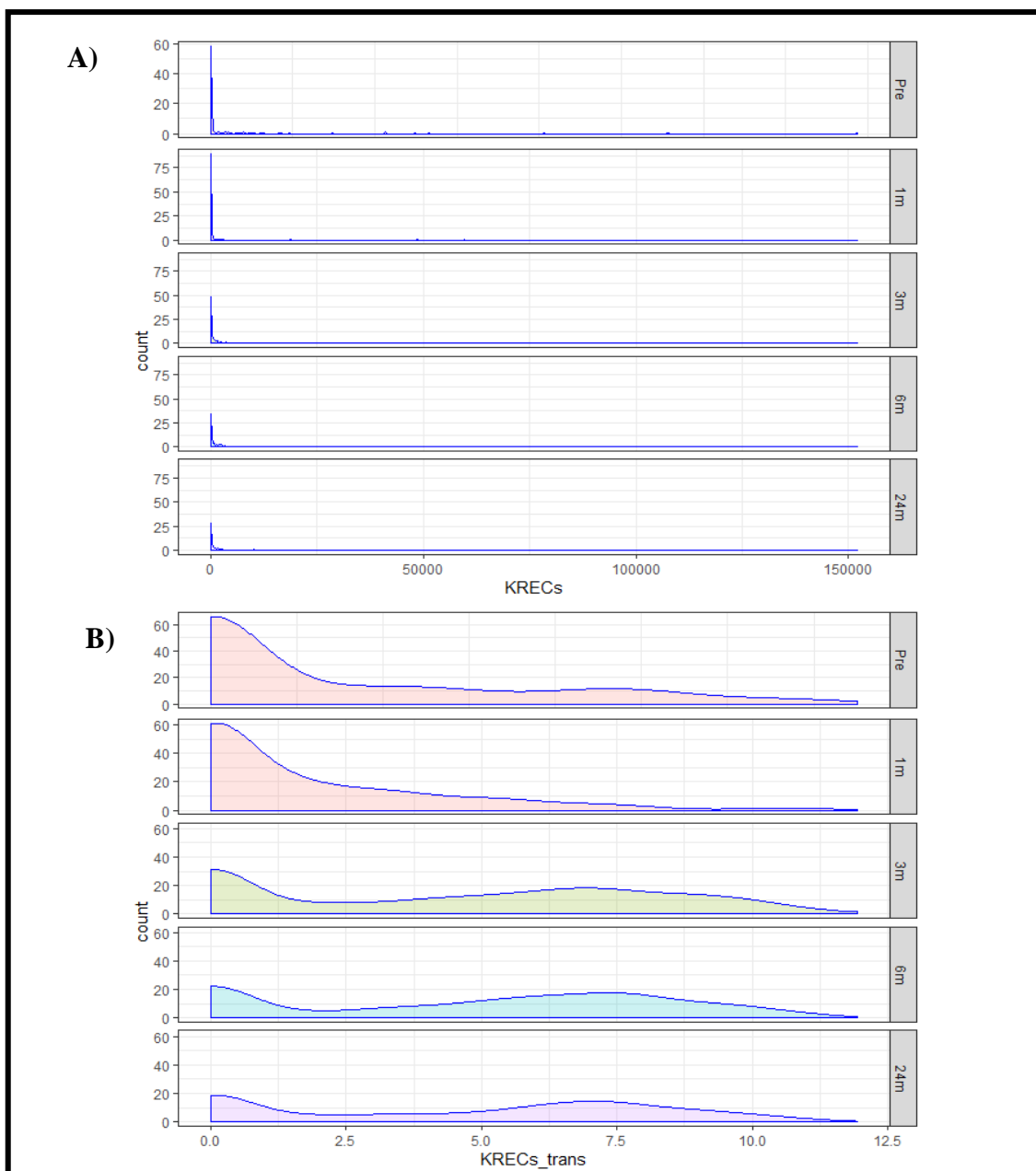


Figura 3. Gráficos de densidad de los valores de la variable KRECs en crudo (A) y tras transformación logarítmica $\log(1+KRECs)$ (B), en los 5 períodos en los que se tomaron medidas.

VARIABLE	PRE-TPH MEDIANA (Q1; Q3)	1M POST TPH MEDIANA (Q1; Q3)	3M POST TPH MEDIANA (Q1; Q3)	6M POST TPH MEDIANA (Q1; Q3)	24M POST TPH MEDIANA (Q1; Q3)
KRECS	0 (0; 22,75)	0.11 (0; 14.21)	118 (0; 1615)	244 (0; 2250)	178 (0; 1825)
LOG (1+KRECS)	0 (0; 3.17)	0.09 (0; 2.72)	4.78 (0; 7.39)	5.50 (0; 7.72)	5.19 (0; 7.51)

Tabla 4. Análisis descriptivo de la tendencia longitudinal de KRECs y $\log(1+KRECs)$ a lo largo de los 5 períodos estudiados, refiriendo la mediana de la distribución en cada medida, así como los valores de Q1 y Q3 que acotan el 50% de los valores.

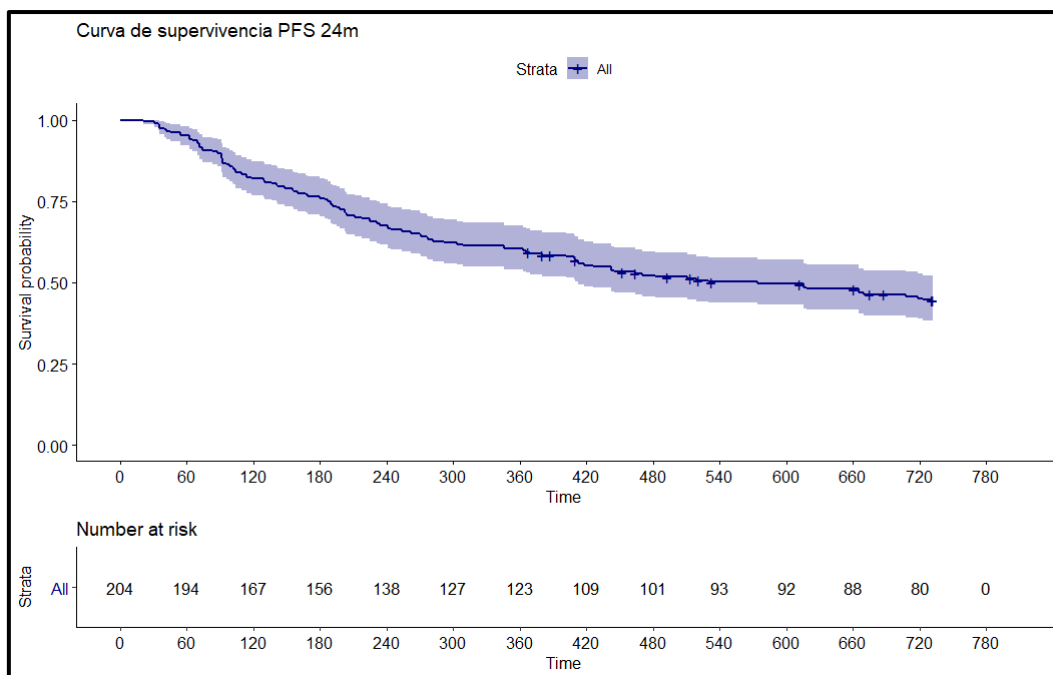


Figura 4. Curva de Kaplan-Meier de supervivencia de los 204 pacientes para el evento PFS a los 2 años.

EVENTO	N (%)	1M (%)	3M (%)	6M (%)	24 (%)
PFS	111 (54.41%)	2 (1.80%)	25 (22.25%)	22 (19.82%)	62 (55.86%)
RECAIDA	52 (25.49%)	1 (1.92%)	15 (28.85%)	9 (17.31%)	27 (51.92%)
NRM	59 (29.92%)	1 (1.79%)	10 (16.95%)	13 (22.03%)	35 (59.32%)

Tabla 5. Incidencia post trasplante de progresión libre de enfermedad (PFS), recaída y muerte sin recaída (NRM) en los 204 pacientes incluidos en el estudio, en cada uno de los 4 periodos definidos. Se refieren las frecuencias absolutas, y entre paréntesis las frecuencias relativas sobre el total de cada tipo de evento.

A los dos años del HSCT, 111 pacientes (54.41%) recaen o fallecen, alcanzándose la mediana de tiempo hasta progresión de la enfermedad a los 574 días. El grueso de los eventos tiene lugar entre los 6 y 24 meses post trasplante, observándose una incidencia de recaída y/o fallecimiento muy baja al mes del trasplante, así como una incidencia de muertes sin recaída ligeramente mayor entre los 3 y 6 meses tras el trasplante y, sin embargo, una mayor incidencia de recaídas entre el mes y los 3 meses post trasplante.

La relación entre los biomarcadores longitudinales TRECs y KRECs y estos eventos se describe gráficamente como se muestra en la Figura 7.

Se observa un descenso en el nivel cuantificado de TRECs entre antes del trasplante y después de este, tanto en los pacientes que acaban recayendo y/o falleciendo como en los

censurados. Dentro de los 3 primeros meses tras el trasplante, tanto los que desarrollan evento como los que no presentan una expresión baja del biomarcador tanto para recaída como para fallecimiento, dándose una variabilidad más elevada entre los pacientes censurados. Bajos niveles de TRECs entre los 3 meses y los 24 meses tras el trasplante parecen asociarse con un peor pronóstico de PFS, intervalo en el que los pacientes censurados si experimentan un incremento en los valores medidos.

En cuanto al biomarcador KRECs, no tiene lugar ese descenso de niveles medidos entre pre/post trasplante que si se observa en el biomarcador TRECs. En este caso, el incremento de expresión del biomarcador tiene lugar entre 1 y 3 meses tras el HSCT, manteniéndose luego estable hasta los 24 meses en los pacientes censurados, y experimentando un descenso en los niveles entre los 6 y 24 meses en los pacientes que fallecen sin recaída. Sí que parece asociarse, sobre todo en el fallecimiento sin recaída, niveles inferiores de KRECs con un peor pronóstico para fallecimiento. Esta asociación no parece apreciarse tan evidente en la recaída, donde la expresión de los KRECs en ambos grupos es bastante similar, siendo notablemente más homogénea la distribución de valores en los pacientes que recaen. Por lo tanto, los niveles más bajos de KRECs en los pacientes que presentan PFS a los 24 meses parecen deberse más al efecto sobre el fallecimiento, que sobre la recaída.

4.2 Modelos de efectos mixtos para cada respuesta longitudinal.

Para estimar los modelos longitudinales, se definió una base de datos longitudinal para cada biomarcador con 738 observaciones totales. Sin embargo, en muchas visitas había valores faltantes, que se distribuían entre las 4 medidas como se muestra en la Tabla 6, y fueron excluidas manteniéndose un total de 645 observaciones para TRECs y 645 observaciones para KRECs.

VARIABLE	1M	3M (%)	6M (%)	24M (%)
	N/TOTAL (%)	N/TOTAL (%)	N/TOTAL (%)	N/TOTAL (%)
TRECS	6/204 (2.94 %)	19/202 (9.45 %)	29/177 (16.38 %)	39/155 (25.16 %)
KRECS	6/204 (2.94 %)	19/202 (9.45 %)	28/177 (15.82 %)	40/155 (25.80 %)

Tabla 6. Distribución de los valores *missing* para las variables TRECs y KRECs en los 4 periodos post trasplante. Se refieren las frecuencias absolutas y, entre paréntesis, las frecuencias relativas sobre el total de valores en cada periodo.

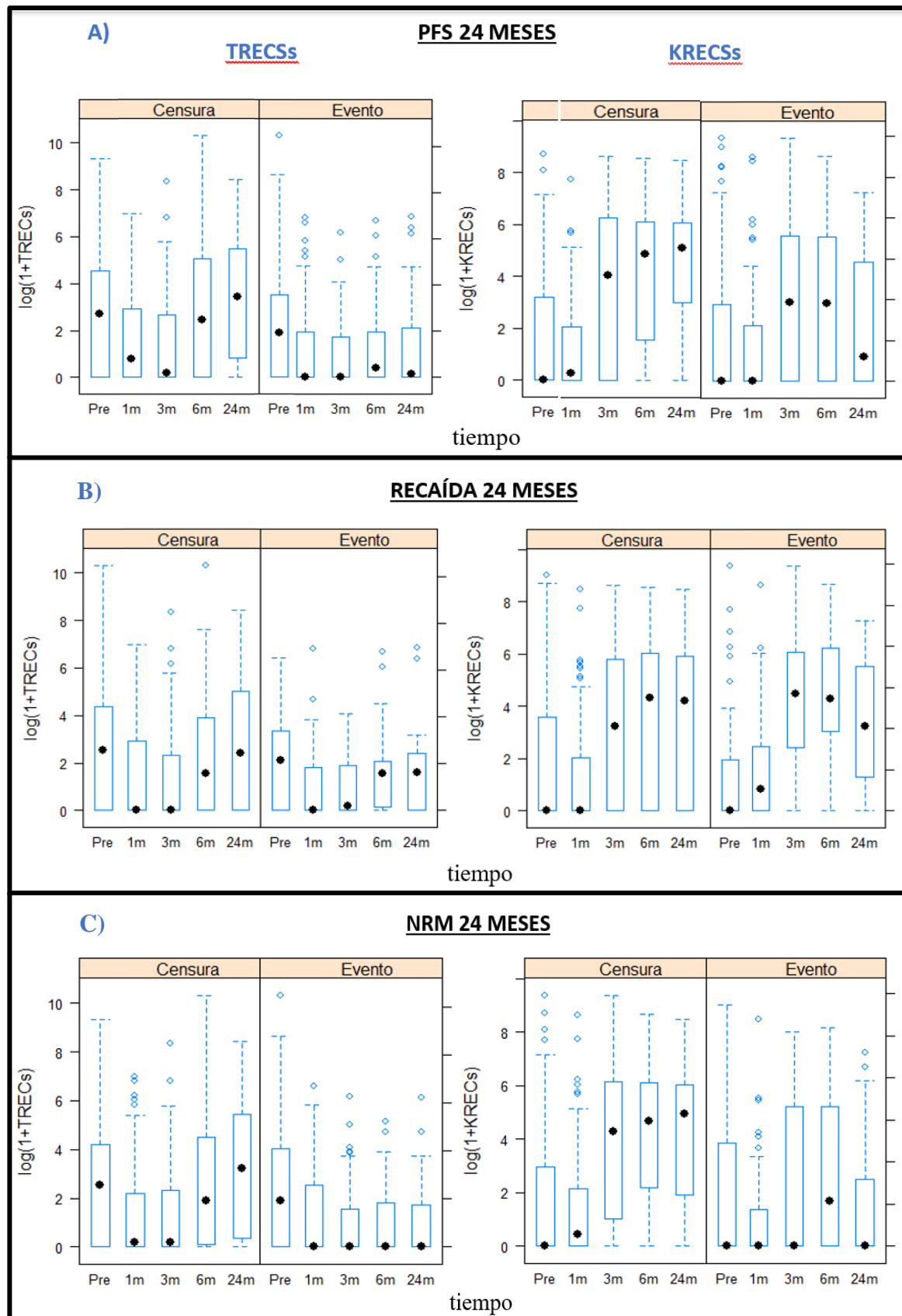


Figura 5. Gráficos de cajas representando la dinámica longitudinal y la variabilidad de los niveles transformados TRECs y KRECs para cada desenlace clínico, estratificado por desarrollo o no del evento específico en cada medida post trasplante.

De cara a definir la estructura de efectos aleatorios, se analizó la dinámica individual de los TRECs y KRECs en cada paciente, con el objetivo de determinar si era necesario incluir únicamente un componente de correlación asumiendo término independiente aleatorio, o también incluir un componente de correlación definiendo la tendencia aleatoria de las variables TRECs y KRECs a lo largo del tiempo (pendiente aleatoria). En la Figura 6 se representa mediante un gráfico de perfil la tendencia de los valores de TRECs post trasplante para 5 individuos aleatorios de la serie de casos analizada. Para ambos biomarcadores se observa una tendencia heterogénea, comportándose de manera diferente en unos pacientes u otros, lo que indica la necesidad de incluir un componente aleatorio para la pendiente, que se contemple posteriormente en la estructura de efectos aleatorios compartidos que define la asociación entre el modelo mixto y el modelo de supervivencia en los modelos *joint*.

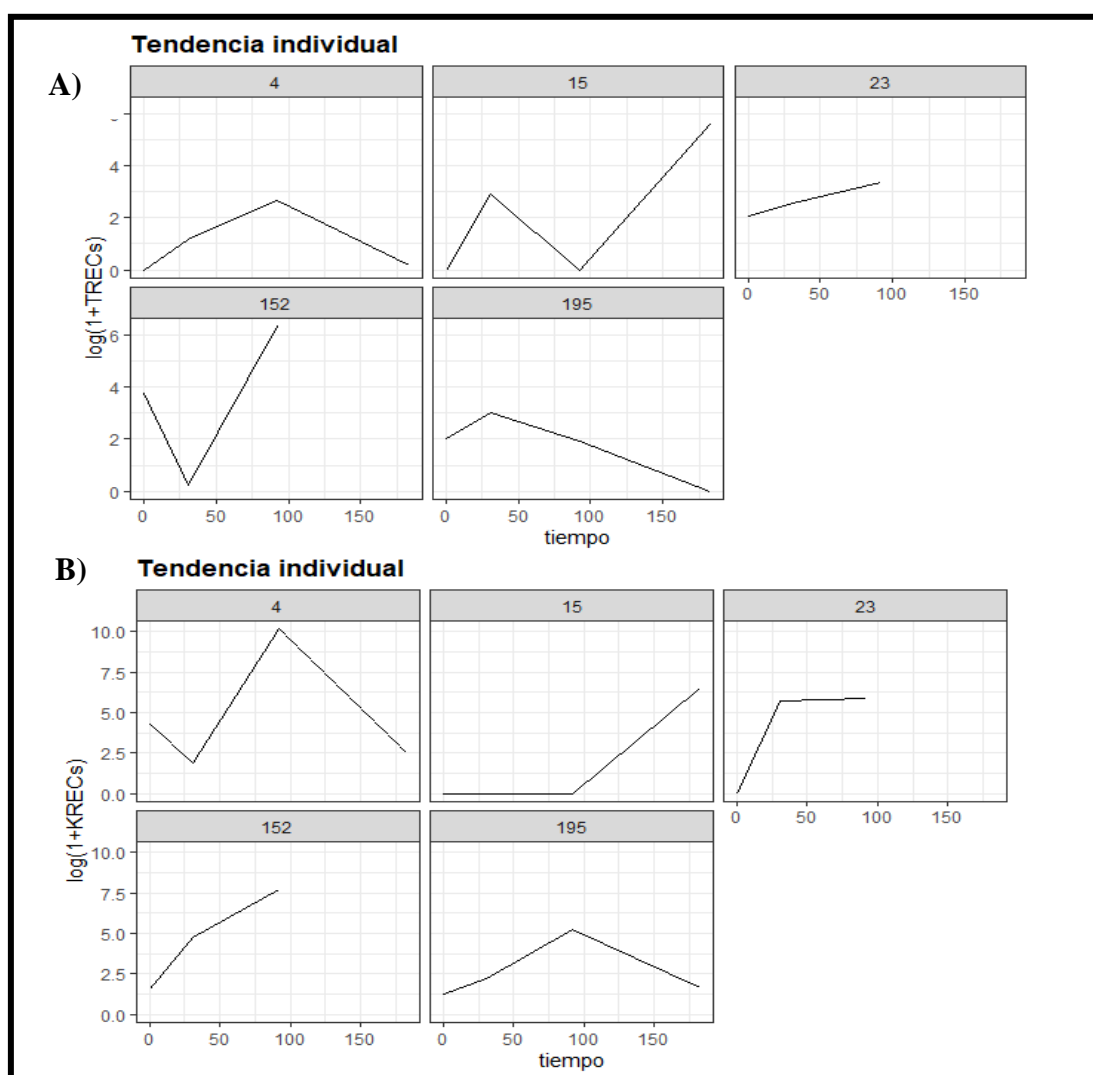


Figura 6. Tendencias longitudinales individuales de los niveles transformados de TRECs (A) y KRECs (B) para 5 pacientes seleccionados aleatoriamente del total incluidos en el estudio.

La metodología aplicada para la estimación de cada modelo multivariante por separado implicó un primer análisis univariante, en el que se seleccionaron aquellas covariables con un p valor inferior a 0.2. La estructura de efectos aleatorios incluyó por lo tanto término independiente aleatorio, definido por la variable identificadora de cada paciente, y un término de pendiente aleatoria definido por la variable tiempo que define las medidas igualmente espaciadas post trasplante. Siguiendo el criterio clínico, la variable que recoge los valores transformados de la medida pre trasplante de TRECs, así como la de KRECs, se mantuvieron en el modelo a expensas de la significación estadística estimada como variable de ajuste fija.

Conociendo que las variables respuesta no siguen un comportamiento normal, se realizó el diagnóstico de los modelos mixtos definitivos para comprobar la hipótesis de normalidad de los residuos estandarizados de Pearson, tanto para los efectos fijos como aleatorios, comprobándose que los términos de error cumplían las hipótesis de normalidad.

Modelos para TRECs

<i>Variable</i>	<i>p valor</i>	<i>Selección</i>
SEXO	0.07	Sí
EDAD	0.69	No
TIPO DE DONANTE	0.12	Sí
PATOLOGÍA HEMATOLÓGICA	0.02	Sí
ACONDICIONAMIENTO	<0.001	Sí
EICH	0.09	Sí
Log (1+TRECs pre TPH)	0.20	Sí
TIEMPO	<0.001	Sí

Tabla 7. Modelos mixtos univariantes para la selección de covariables fijas para TRECs.

Mediante el análisis univariante, se seleccionan todas las covariables de ajuste incluidas en el análisis a excepción de la edad. Se procedió a modelizar un modelo multivariante que incluyese las covariables estimadas estadísticamente significativas ($p < 0.05$), siguiendo un método de selección de covariables hacia atrás (*backward*) a partir del modelo máximo. Los modelos anidados se compararon mediante un análisis de varianza,

y para modelos no anidados se utilizaron para la selección de modelos los criterios de información de Akaike (AIC) y bayesiano (BIC):


<i>Modelo máximo</i>	<i>AIC</i>	<i>BIC</i>
$t1_mmax \leftarrow lme(\log(1+TRECs) \sim \text{Sexo} + EICH + \text{Acondicionamiento} + \text{Tipo_donante} + \text{Patología_hematológica} + \log(1+TRECs \text{ preTPH}) + \text{Tiempo}, \text{random} = \sim \text{Tiempo} \text{id}, \text{data} = \text{base_sm_IT})$	2722.50	2775.98
Selección hacia atrás de covariables 		
<i>Modelo definitivo</i>	<i>AIC</i>	<i>BIC</i>
$t1_def \leftarrow lme(\log(1+TRECs) \sim \text{Tipo_donante} + \log(1+TRECs \text{ preTPH}) + \text{Tiempo}, \text{random} = \sim \text{Tiempo} \text{id}, \text{data} = \text{base_sm_IT})$	2719.50	2764.09

Tabla 8. Transición desde el modelo multivariante máximo hasta el modelo multivariante definitivo para TRECs.

El modelo multivariante de efectos mixtos para TRECs incluye en el componente de efectos fijos las covariables: tipo de donante en el trasplante, la medida transformada pre trasplante de TRECs y la variable de tiempo que define las medidas igualmente espaciadas. Se compararon los resultados obtenidos asumiendo una estructura de efectos aleatorios con pendiente y término independiente aleatorios, con modelos cuyo componente de efectos aleatorios incluía únicamente un *intercept* aleatorio, comprobándose que el modelo que incluye la pendiente aleatoria presenta un mejor ajuste. Se comprobaron también las interacciones entre el tiempo y las covariables tipo de donante y medida transformada pre trasplante de TRECs, no estimándose como estadísticamente significativa ninguna de las dos, por lo que se mantuvo este modelo como el modelo definitivo.

Proceso longitudinal $\log(1+TRECs)$				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p val</i>	<i>anova p val</i>
Tipo Donante				
Donante DNE	Cat. Ref.	-	-	<0.001
Donante HAPLO	-0.957	0.239	0.051	
Donante HLAid	0.013	0.490	0.959	
Donante TSCU	-1.032	0.257	<0.001	
Log(1+TRECs_preTPH)	0.055	0.042	0.199	0.199
Tiempo	0.007	0.001	<0.001	<0.001

Tabla 9. Resultados del modelo multivariante de efectos mixtos para la variable respuesta $\log(1+TRECs)$. Cat. Ref: Categoría de referencia

Modelos para KRECs

<i>Variable</i>	<i>p valor</i>	<i>Selección</i>
SEXO	0.57	No
EDAD	0.86	No
TIPO DE DONANTE	0.11	Sí
PATOLOGÍA HEMATOLÓGICA	0.04	Sí
ACONDICIONAMIENTO	0.35	No
EICH	0.41	No
Log (1+KRECs pre TPH)	0.23	Sí
TIEMPO	<0.001	Sí

Tabla 9. Modelos mixtos univariantes para la selección de covariables fijas para KRECs

Mediante el análisis univariante se seleccionaron para incluir como efectos fijos en modelo máximo las covariables tipo de donante, patología hematológica, y la variable de tiempo.


<i>Modelo máximo</i>	<i>AIC</i>	<i>BIC</i>
$k1_mmax \leftarrow lme(\log(1+KRECs) \sim \text{Tipo_donante} + \text{Patología_hematológica} + \log(1+KRECs \text{ preTPH}) + \text{Tiempo}, \text{random} = \sim \text{Tiempo} \text{id}, \text{data} = \text{base_sm_IK}, \text{control} = lmeControl(\text{opt} = "optim"))$	3388.36	3450.71
		
<i>Modelo definitivo</i>	<i>AIC</i>	<i>BIC</i>
$k1_def \leftarrow lme(\log(1+KRECs) \sim \text{Patología_hematológica} + \log(1+KRECs \text{ preTPH}) + \text{Tiempo}, \text{random} = \sim \text{Tiempo} \text{id}, \text{data} = \text{base_sm_IK}, \text{control} = lmeControl(\text{opt} = "optim"))$	3385.78	3434.83

Tabla 10. Transición desde el modelo multivariante máximo hasta el modelo multivariante definitivo para KRECs.

En el modelo definitivo se mantuvieron como variables fijas la hemopatía del paciente, la medida transformada pre trasplante y la variable de tiempo. Se comprobaron las interacciones de las variables fijas con el tiempo, no estimándose como significativa ninguna de las dos, por lo que se mantuvo el modelo mostrado como el modelo longitudinal definitivo para el biomarcador KRECs.

Proceso longitudinal log(1+KRECs)				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p val</i>	<i>anova p val</i>
Hemopatía previa				
Hemopatía LA/SMD	Cat.ref	-	-	0.041
Hemopatía Aplasia	-0.007	2.065	0.997	
Hemopatía MM	-0.425	0.633	0.503	
Otras hemopatías	-0.942	0.737	0.203	
Hemopatía SLP	-1.297	0.415	0.002	
Log(1+KRECs_preTPH)	0.083	0.059	0.163	0.163
Tiempo	0.013	0.002	<0.001	<0.001

Tabla 11. Resultados del modelo multivariante de efectos mixtos para la variable respuesta log(1+KRECs). Cat. Ref: Categoría de referencia

4.3 Modelos de tiempo hasta el evento para cada biomarcador.

El otro elemento necesario para modelizar los modelos *joint* son los modelos de análisis de supervivencia, que van a estimar el efecto de determinadas covariables sobre el tiempo hasta el evento, en este caso el tiempo libre de progresión, mediante el modelo de riesgos proporcionales de Cox. Para ello se define una base de datos en formato ancho con una única observación por paciente (204 observaciones totales), recogiendo en la variable *tiseg* el tiempo hasta recaída, fallecimiento o censura, censurándose a los 732 días si en ese punto no se ha observado la aparición del evento. La metodología aplicada para estimar los modelos definitivos es similar a la aplicada en la modelización de la respuesta longitudinal, con un primer paso de análisis univariante seleccionando aquellas variables con un p valor menor a 0.2, y posteriormente realizando una selección de covariables hacía atrás ($p < 0.05$) a partir del modelo máximo.

Modelos para TRECs

<i>Variable</i>	<i>p valor</i>	<i>Selección</i>
SEXO	0.11	Sí
EDAD	0.22	No
TIPO DE DONANTE	0.13	Sí
PATOLOGÍA HEMATOLÓGICA	0.68	No
ACONDICIONAMIENTO	0.67	No
EICH	0.20	No
Log (1+TRECs pre TPH)	0.03	Sí

Tabla 12. Modelos de Cox univariantes para la selección de covariables fijas para PFS

En este primer paso, se seleccionan las covariables sexo, tipo de donante al trasplante y la medida transformada pre trasplante de TRECs. A partir del modelo máximo, se van excluyendo aquellas variables que no alcanzan la significación estadística.


<i>Modelo máximo</i>	<i>AIC</i>	<i>BIC</i>
coxt_max \leftarrow coxph (Surv (tiseq, PFS) ~ Sexo + Tipo_donante+ log (1+TRECs preTPH), data = baseimp_wideTREC)	1098.04	1111.59
		
<i>Modelo definitivo</i>	<i>AIC</i>	<i>BIC</i>
coxt_def \leftarrow coxph (Surv (tiseq, PFS) ~ (1+TRECs preTPH) , data = baseimp_wideTREC)	1098.55	1101.26

Tabla 13. Transición desde el modelo multivariante de Cox máximo hasta el modelo multivariante definitivo para TRECs con variable respuesta PFS

Proceso tiempo hasta PFS					
<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IC HR 95%</i>
Log(1+TRECs_preTPH)	-0.096	0.908	0.045	0.031	0.833, 0.991

Tabla 14. Resultados del modelo de Cox definitivo para TRECs con variable respuesta PFS

El modelo definitivo incluye únicamente la covariable transformada medida pre trasplante de TRECs. No se consideraron interacciones importantes desde el punto de vista clínico, por lo que no se comprobaron posibles interacciones con otras covariables. No se observa asociación entre las variables sexo y tipo de donante al trasplante con la PFS, pero sí con los niveles transformados pre trasplante de TRECs. El modelo estima un *hazard ratio* (HR) para la medida pre TPH de 0.91 aproximadamente, indicando que, dados dos pacientes, el incremento relativo de una unidad en los niveles cuantificados de log(1+TRECs) antes del trasplante alogénico reduce en un 9.1% el riesgo de recaer o fallecer a los 24 meses tras el trasplante.

Modelos para KRECs

Al tratarse de los mismos pacientes en la base ancha para KRECs que para TRECs, el único modelo diferente a lo estimado en los modelos de Cox univariantes para TRECs es el modelo que incluye como covariable la medida pre trasplante de KRECs, que se mantendrá en el modelo multivariante a pesar de la significación estadística que se le estime. Aun sabiendo que las covariables seleccionadas en el primer paso fueron

excluidas del modelo multivariante para el biomarcador TRECs, se comprueba en este caso por si al ajustar por la medida pre trasplante de KRECs alguna de ellas alcanzase la significación estadística.


<i>Modelo máximo</i>	<i>AIC</i>	<i>BIC</i>
coxk_max \leftarrow coxph (Surv (tiseq, PFS) ~ Sexo + Tipo_donante+ log (1+KRECs preTPH), data = baseimp_wideKREC)	1102.56	1116.11
		
<i>Modelo definitivo</i>	<i>AIC</i>	<i>BIC</i>
coxk_def \leftarrow coxph (Surv (tiseq, PFS) ~ log (1+KRECs preTPH), data = baseimp_wideKREC)	1103.20	1105.91

Tabla 15. Transición desde el modelo multivariante de Cox máximo hasta el modelo multivariante definitivo para KRECs con variable respuesta PFS

Finalmente, el modelo de tiempo hasta el evento definitivo únicamente se ajusta por los niveles transformados pre trasplante de KRECs, al igual que ocurría para el biomarcador TRECs. En este caso, sin embargo, no se estima como estadísticamente significativa la asociación de esta covariable sobre el riesgo de PFS, no encontrándose evidencia suficiente de que el tiempo hasta progresión se vea modificado por los niveles pre trasplante de KRECs.

Proceso tiempo hasta PFS					
<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IC HR 95%</i>
Log(1+KRECs_preTPH)	-0.016	0.984	0.038	0.674	0.913, 1.060

Tabla 16. Resultados del modelo de Cox definitivo para KRECs con variable respuesta PFS

4.4 Modelos *joint* para cada biomarcador

Tras estimar el modelo mixto para la respuesta longitudinal de cada biomarcador, así como el modelo de regresión que define el riesgo sobre el tiempo hasta el evento, se puede proceder a la elaboración de los modelos *joint* que integrará la información aportada por cada uno de ellos para cuantificar la asociación entre el componente longitudinal y el tiempo hasta progresión de la enfermedad. Como se mencionó previamente (ver Epígrafe 3.2.5), se opta por trabajar siguiendo un enfoque bayesiano en la modelización, asumiendo una distribución a priori no informativa, otorgando todo el peso sobre la distribución a posteriori a los datos muestrales.

Para la programación de los modelos en *R*, se utilizó el paquete desarrollados por Dimitris Rizopoulos *JMbayes* ^[14]. Este último, que incluye el comando *jointmodelBayes* utilizado para la elaboración de los modelos, estima los modelos por defecto mediante 20000 iteraciones, con un tamaño de *burning* de 3000 iteraciones iniciales, y un intervalo de *thinning* seleccionando cada 10 valores simulados. De forma inicial se estimaron los modelos con este patrón de control, pero tras analizar el diagnóstico de convergencia de las cadenas, así como comparando los modelos mediante los criterios de información de *deviance* (DIC) y la media a posteriori de la *deviance* (pD), se pudo ver que no parecía haberse alcanzado la distribución estacionaria correctamente, sobre todo apreciándose una elevada autocorrelación en los valores. Para tratar de solventar este problema, se incrementó el número de iteraciones a 60000 (*n.iter*), quemando las primeras 10000 iteraciones (*n.burnin*), y seleccionando cada 30 valores simulados (*n.thin*). Se comprobó la convergencia de las cadenas en los modelos mediante gráficos de trazas, autocorrelación y cuantiles.

Modelo joint para TRECs

Modelo mixto longitudinal

```
t1_def <- lme (log (1+TRECs) ~ Tipo_donante + log (1+TRECs preTPH) +  
Tiempo, random = ~ Tiempo | id, data = base_sm_IT
```

Modelo de Cox

```
coxt_ded <- coxph (Surv (tiseq, PFS) ~ (1+TRECs preTPH) ,  
data = baseimp_wideTREC)
```

Modelo Joint

```
joint_T<-jointModelBayes(t1_def, coxt_def, timeVar = "Tiempo",  
control = list(n.iter=60000,n.burnin=10000,n.thin=30))
```

Modelo *joint* para TRECs. Asociación entre dinámica de log(1+TRECs) y tiempo hasta PFS

Tras completarse el proceso iterativo, el programa devuelve dos resultados, uno correspondiente al proceso longitudinal para las medidas repetidas del biomarcador y otro para el proceso de tiempo hasta PFS, que incluye el parámetro de asociación que define esa relación entre la respuesta longitudinal y el riesgo sobre PFS. Dicho coeficiente es similar al coeficiente estimado en el modelo de Cox y, aplicando una transformación exponencial, se define el *hazard ratio* sobre la respuesta de supervivencia.

Proceso longitudinal log(1+TRECs)				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
Tipo Donante				
Donante DNE	Cat. ref	-	-	-
Donante HAPLO	-0.875	0.013	0.046	-1.824, -0.023
Donante HLAid	0.049	0.019	0.829	-0.425, 0.525
Donante TSCU	-0.879	0.010	0.003	-1.495, -0.329
Log(1+TRECs_preTPH)	0.053	0.002	0.268	-0.043, 0.147
Tiempo	0.005	0.001	<0.001	0.003, 0.008

Tabla 17. Resultados del proceso longitudinal estimado mediante la metodología *joint* para log(1+TRECs)

Proceso tiempo hasta PFS					
<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP HR 95%</i>
Log(1+TRECs_preTPH)	-0.046	0.956	0.003	0.461	0.849, 1.079
Assoct log(1+TRECs)	-1.130	0.323	0.044	<0.001	0.171, 0.550

Tabla 18. Resultados del proceso de supervivencia estimado mediante la metodología *joint* para PFS y log(1+TRECs). Assoct: Asociación entre proceso longitudinal y de supervivencia.

Los resultados estimados para el modelo longitudinal no se ven prácticamente modificados en comparación con el modelo mixto multivariante clásico, con la única diferencia de que el tipo de donante haploidéntico es estimado estadísticamente significativo sobre la respuesta longitudinal logarítmica de TRECs frente a la categoría donante DNE de referencia. El proceso de tiempo hasta PFS sí que se ve modificado ya que, al integrar la información de las medidas repetidas del biomarcador, la covariable fija de la medida transformada pre trasplante de TRECs, que en el modelo de Cox se estimó que se asociaba de forma significativa sobre el riesgo de recaer o fallecer (HR: 0.908 [0.8327, 0.9914]), deja de tener efecto significativo.

La dinámica longitudinal de los TRECs sí que se asocia de forma significativa con el tiempo libre de progresión. El HR a posteriori estimado tiene un valor de 0.323 [IP95%: 0.171; 0.55], estando ajustado tanto por las covariables fijas incluidas en el modelo longitudinal, como por las incluidas en el modelo de Cox. Dados dos pacientes, con características similares en cuanto a tipo de donante para el HSCT, niveles pre trasplante de log(1+TRECs), y misma medida post trasplante, se estima que el incremento de una unidad en el log(1+TRECs) en uno de ellos respecto al otro se asocia con una reducción

del 67.7% [44.9%; 82.9%] del riesgo de recaer o fallecer a los 24 meses post trasplante para ese paciente. Por lo tanto, los niveles del biomarcador TRECs de forma individual se estiman como factor protector sobre el tiempo libre de progresión tras el trasplante alogénico.

Modelo joint para KRECs

Modelo mixto longitudinal

```
k1_def <- lme (log (1+KRECs) ~ Patología_hematológica + log (1+KRECs preTPH) + Tiempo,
random = ~ Tiempo| id , data = base_sm_1K,
control = lmeControl(opt = "optim"))
```

Modelo de Cox

```
oxk_def<- coxph (Surv (tiseq, PFS) ~ log (1+KRECs preTPH), data = baseimp_wideKREC)
```

Modelo Joint

```
joint_K<-jointModelBayes(k1_def, coxk_def, timeVar = "Tiempo",
control = list(n.iter=60000,n.burnin=10000,n.thin=30))
```

Modelo *joint* para KRECs. Asociación entre dinámica de log(1+KRECs) y tiempo hasta PFS

Proceso longitudinal log(1+KRECs)				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
Hemopatía previa				
Hemopatía LA/SMD	Cat.ref.	-	-	-
Hemopatía Aplasia	0.631	0.099	0.798	-4.124, 5.747
Hemopatía MM	-0.1185	0.033	0.872	-1.630, 1.393
Otras hemopatías	-0.897	0.041	0.357	-2.843, 1.067
Hemopatía SLP	-0.846	0.024	0.124	-1.923, 0.219
Log(1+KRECs_preTPH)	0.080	0.0036	0.347	-0.084, 0.244
Tiempo	0.012	0.0001	<0.001	0.007, 0.016

Tabla 19. Resultados del proceso longitudinal estimado mediante la metodología *joint* para log(1+KRECs)

Proceso tiempo hasta PFS					
<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP HR 95%</i>
Log(1+KRECs_preTPH)	-0.001	0.999	0.001	0.986	0.919, 1.084
Assoct log(1+KRECs)	-0.280	0.756	0.010	0.011	0.614, 0.925

Tabla 20. Resultados del proceso de supervivencia estimado mediante la metodología *joint* para PFS y log(1+KRECs). Assoct: Asociación entre proceso longitudinal y de supervivencia.

Los resultados del modelo *joint* para el biomarcador KRECs permiten también inferir un valor pronóstico de dicho biomarcador sobre el tiempo hasta recaída o muerte a los 2 años post TPH. A diferencia de lo que ocurría en el proceso longitudinal del modelo conjunto para TRECs, al integrar ambos procesos sí que se aprecia una modificación relevante de los coeficientes estimados para la covariable patología hematológica, perdiéndose la asociación estadísticamente significativa entre la categoría síndrome linfoproliferativo (SLP), frente a la categoría LA/SMD de referencia, con los niveles de $\log(1+\text{KRECs})$. El proceso de tiempo hasta evento no se ve modificado ya que no se estimó en el modelo de Cox un efecto significativo de la medida pre trasplante de KRECs sobre el riesgo de PFS. El resultado fundamental e interesante que deriva de la integración simultánea de la información de ambos procesos es la asociación estadísticamente significativa estimada entre el tiempo libre de progresión y la respuesta logarítmica longitudinal de KRECs, con un HR estimado de 0.76 [IP95%: 0.61;0.92]. Este resultado indica que, para dos pacientes con misma patología hematológica, niveles pre trasplante de $\log(1+\text{KRECs})$ y en el mismo periodo post TPH, un aumento de una unidad en el $\log(1+\text{KRECs})$ tras el trasplante en uno respecto al otro conlleva una reducción del riesgo sobre recaída o muerte a los 24 meses del 24.4% [7.5%; 38.6%].

Por lo tanto, analizando el valor pronóstico tras el trasplante de la respuesta longitudinal de cada biomarcador de forma independiente, ambas se asocian de manera significativa con el tiempo libre de progresión, asociándose a un menor riesgo de recaída o muerte para un paciente conforme se incrementa su expresión. Esta asociación es mayor con los niveles transformados de TRECs que los de KRECs, asociados a una reducción mayor del riesgo sobre PFS, aunque sin embargo hay que tener en cuenta la mayor dispersión de valores de KRECs que de TRECs, que conlleva que los niveles transformados de KRECs suelen ser más elevados y, por ende, incrementar una unidad tenga menor impacto sobre el riesgo, a que se produzca un aumento en los niveles de TRECs.

4.5 Modelos *joint* para respuesta longitudinal multivariante

En la práctica clínica, conocer el valor pronóstico de cada uno de los biomarcadores por separado es interesante y puede tener sus implicaciones, sin embargo, lo habitual es trabajar con el valor de ambos biomarcadores y, por lo tanto, el interés reside en conocer cómo se asocia el comportamiento longitudinal de ambos biomarcadores

simultáneamente con el tiempo libre de progresión a los 24 meses tras el trasplante alogénico.

Para ello, se elabora un modelo *joint* multivariante que incluye dos variables respuesta, integrando la información de ambas dinámicas longitudinales con la información aportada por el modelo de supervivencia, y devolviendo un resultado ajustado por y para ambos biomarcadores.

Es necesario definir una base de datos longitudinal con los datos de las dos variables de interés, así como el resto de covariables seleccionadas en los procesos univariantes, y una base en formato ancho que incluya las medidas pre trasplante de cada biomarcador. Para estimar un modelo mixto multivariante con dos variables respuesta se utiliza la función *mvglmer* del paquete *JMBayes* ^[14], que estima los modelos mixtos de medidas repetidas para cada variable respuesta mediante métodos bayesianos. Para estimar el modelo *joint* multivariante se utiliza la función *mvJointModelBayes* del mismo paquete.

Se asume un comportamiento gaussiano para las respuestas longitudinales de TRECs y KRECs, aún a sabiendas de que no es la distribución que mejor refleja el comportamiento de estas variables de interés. Se trató de estimar un modelo *joint* multivariante partiendo de un modelo mixto multivariante en el que se asumía una distribución gamma para ambas variables, ya que se determinó que es la distribución que mejor representa dicho comportamiento, pero por limitaciones del software no se pudo incluir una función que definiese los parámetros de escala y forma asociados a la distribución gamma de cada biomarcador, no pudiendo procesarse los modelos.

Modelo mixto multivariante

```
MultiTK<-mvglmer(list(log(1+TRECs) ~ Tipo_donante + log(1+TRECs preTPH) +  
Tiempo +(Tiempo | id), log(1+KRECs) ~ Patología_hematológica + log(1+KRECs preTPH)  
+ Tiempo + (Tiempo | id)), data = base_MULTI, families = list(gaussian, gaussian))
```

Modelo bivariante Cox

```
cox_multiTK <- coxph(Surv(tiseg, PFS) ~ log(1+TRECs preTPH) + log(1+KRECs  
preTPH), data = base_multicox)
```

Modelo Joint multivariante

```
JBmulti <- mvJointModelBayes(MultiKT,cox_multi,timeVar = "Tiempo",  
control = list(n.iter=60000,n.burnin=10000,n.thin=30)))
```

Modelo *joint* multivariante para ambas variables respuesta $\log(1+\text{TRECs})$ y $\log(1+\text{KRECs})$.
Asociación entre dinámica de ambas y el tiempo hasta PFS

A)	Proceso longitudinal: Respuesta log(1+TRECs)				
	<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
	<i>Tipo Donante</i>				
	Donante DNE	Cat.ref.	-	-	-
	Donante HAPLO	-0.819	0.016	0.086	-1.785, 0.102
	Donante HLAid	0.066	0.008	0.786	-0.428, 0.560
	Donante TSCU	-0.986	0.009	<0.001	-1.541, -0.441
	Log(1+TRECs_preTPH)	0.062	0.001	0.144	-0.019, 0.142
	Tiempo	0.004	0.001	0.008	0.001, 0.007
B)	Proceso longitudinal: Respuesta log(1+KRECs)				
	<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
	<i>Hemopatía previa</i>				
	Hemopatía LA/SMD	Cat.ref.	-	-	-
	Hemopatía Aplasia	-0.581	0.007	0.748	-4.269, 3.061
	Hemopatía MM	0.100	0.017	0.872	-0.951, 1.180
	Otras hemopatías	-0.265	0.019	0.688	-1.427, 1.001
	Hemopatía SLP	-0.846	0.011	<0.001	-1.804, -0.457
	Log(1+KRECs_preTPH)	0.099	0.001	0.046	0.001, 0.196
	Tiempo	0.013	0.001	<0.001	0.008, 0.018

Tabla 21. Resultados del proceso longitudinal multivariante estimado mediante la metodología *joint* para log(1+TRECs) (A) y log(1+KRECs) (B)

Proceso tiempo hasta PFS					
<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP HR 95%</i>
Log(1+TRECs_preTPH)	-0.111	0.895	0.005	0.136	0.586, 1.077
Log(1+KRECs_preTPH)	-0.029	0.973	0.005	0.946	0.597, 1.115
Assoct log(1+TRECs)	-1.212	0.297	0.130	0.020	0.001, 0.970
Assoct log(1+KRECs)	0.001	1.001	0.019	0.336	0.754, 4.053

Tabla 22. Resultados del proceso multivariante de supervivencia estimado mediante la metodología *joint* para PFS y ambas dinámicas longitudinales. Se estima un coeficiente de asociación ajustado para cada variable respuesta longitudinal.

Al ajustar el modelo conjunto incluyendo ambas respuestas logarítmicas de TRECs y KRECs longitudinales, destaca ante todo la pérdida de asociación significativa entre la respuesta longitudinal transformada de KRECs y el riesgo de tiempo libre de progresión. Los modelos longitudinales ajustados por ambas variables respuesta sufren alguna ligera modificación en comparación con los resultados obtenidos para cada modelo por

separado, entre las que destaca la significación estadística alcanzada por la medida transformada pre trasplante de KRECs.

No obstante, el resultado más interesante son las estimaciones de los parámetros de asociación entre la respuesta longitudinal y el tiempo hasta evento. La asociación entre los niveles en escala logarítmica de KRECs y el tiempo libre de progresión deja de estimarse significativa, estimándose un HR superior a 1 para este biomarcador al ser ajustado por la respuesta de TRECs. Los niveles transformados de TRECs mantienen la asociación estadísticamente significativa con el desenlace clínico estudiado, reduciéndose ligeramente el HR estimado de 0.323 a 0.297. Sin embargo, llama la atención la excesiva amplitud de los intervalos de probabilidad al 95% estimados (IP95% $\log(1+\text{TRECs})$): [0.0005, 0.970]; IP95% $\log(1+\text{KRECs})$: [0.754, 4.053]. Este hecho puede ser debido a la relación o asociación existente entre ambas variables ya que, observando el análisis descriptivo de las variables, parece evidente la correlación entre ambos biomarcadores, incrementando los niveles de TRECs cuando incrementan los de KRECs, y viceversa, lo que deriva en un fenómeno de colinealidad al incluir ambas en el modelo. No se encontró ningún diagnóstico en el software para verificar dicha colinealidad en estos modelos. Habrá que ver si el problema se mantiene al estimar la asociación sobre el riesgo por causa específica cuando se aplique el modelo multiestado sin estados intermedios.

Así, dados dos pacientes con un nivel cuantificado similar de $\log(1+\text{KRECs})$, y ajustado por el resto de covariables incluidas (tipo de donante al trasplante, hemopatía de origen, niveles transformados pre trasplante de TRECs y KRECs y medida post trasplante), se estima que el incremento de una unidad en el $\log(1+\text{TRECs})$ en uno respecto al otro va a resultar en una reducción del 70.2% [3%; 99%] del riesgo de recaer o fallecer a los 24 meses del trasplante. Estos resultados han de ser interpretados con cautela ya que, como se ha mencionado, la escasa precisión de las estimaciones puede deberse a la relación estadística entre ambos biomarcadores.

4.6 Modelos *joint* para riesgos competitivos con modelos multiestado

El desenlace clínico PFS estudiado en los análisis previos se compone de dos eventos específicos, que son la recaída y el fallecimiento sin recaída (NRM). Estas dos causas específicas pueden contemplarse como un proceso de riesgos competitivos, en el que la recaída es el evento principal, y el fallecimiento el evento competitivo que, si tiene lugar

en un tiempo previo, impide observar la recaída. Este proceso se puede enfocar desde una visión multiestado sin eventos intermedios, en el que se parte de un estado inicial (trasplante HSCT alogénico, E1), pudiendo darse dos posibles transiciones finales, del trasplante a recaída (T1) o de trasplante a fallecimiento (T2), estimando así la asociación entre las variables de interés y el tiempo hasta cada causa específica, tanto de forma univariante como multivariante.

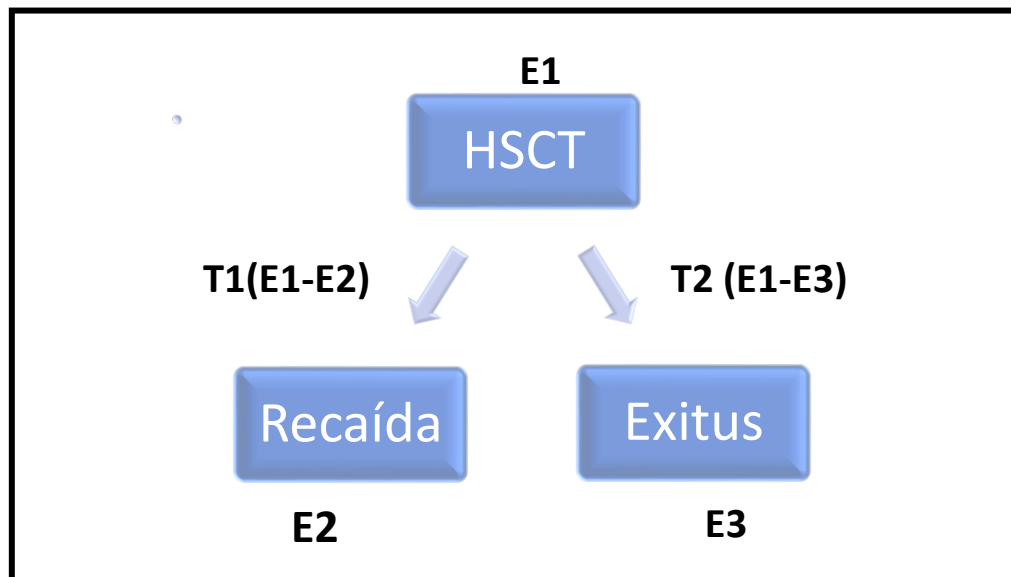


Figura 7. Diagrama esquemático del modelo multiestado aplicado en el estudio. El trasplante HSCT es el estado inicial (E1), con dos transiciones posibles, una transición (T1) a un estado E2 que refiere la recaída del paciente, y otra transición (T2) al fallecimiento (E3). No se contempla la transición entre los estados E2 y E3.

Para su programación en R, es necesario primero definir una base de datos con formato multiestado (una observación por paciente por cada transición posible), utilizando el paquete *mstate* ^[41] y, en especial, la función *msprep* mediante la cual se construye la base de datos mencionada.

Para ello, es necesario definir una matriz de transición, incluyendo las 2 transiciones posibles:

from	to			
	Trasp	Recaída	Muerte	
Trasp	NA	1	2	
Recaída	NA	NA	NA	
Muerte	NA	NA	NA	

Matriz de transición para definir los estados y transiciones posibles.

Y una variable de tiempo hasta cada causa específica, así como una variable de censura para cada una de éstas (0 → Censura, 1 → Evento). Finalmente, mediante la función mencionada, se construye la base de datos adecuada para estimar un modelo multiestado de tiempo hasta el evento.

En cuanto al modelo de efectos mixtos longitudinal que define el componente longitudinal del modelo *joint*, es necesario estimar los modelos mediante la función *mvglmer*, ya que se requiere un modelo estimado bajo métodos bayesianos para estimar el modelo conjunto multiestado.

Para este análisis, se estimó un modelo *joint* univariante para cada respuesta longitudinal por separado, así como un modelo multivariante integrando la información de ambas dinámicas longitudinales.

Modelo *joint* multiestado para TRECs

Para estimar el coeficiente de asociación entre el modelo longitudinal y el modelo multiestado de Cox es necesario incluir en el código del modelo *joint* un componente que defina la interacción entre la respuesta longitudinal y cada causa específica.

Modelo mixto longitudinal

```
Multilong_T ← mvglmer (list (log (1+TRECs) ~ Tipo_donante + log (1+TRECs preTPH) +  
Tiempo + (Tiempo | id), data = base_sm_IT, families = list (gaussian))
```

Modelo multiestado

```
cox_msT <- coxph (Surv(Tstart, Tstop, status) ~ log (TRECs preTPH) + strata(trans)+cluster(id),  
data = mst, x=TRUE, model = TRUE)
```

Modelo *Joint* multiestado

```
interacts <- list("log(1+TRECs)" = ~ strata(trans) - 1)
```

```
JM_msT <- mvJointModelBayes(Multi_longT, cox_msT, timeVar = "Tiempo",  
multiState = TRUE, Interactions = interacts, data_MultiState = mst, idVar_MultiState = "id",  
control = list(equal.strata.knots = TRUE, equal.strata.bound.knots = TRUE))
```

Modelo *joint* multiestado para la respuesta longitudinal log(1+TRECs) y el tiempo hasta cada causa específica.

Los resultados obtenidos para el modelo longitudinal muestran una similitud bastante clara entre la estimación desde el enfoque frecuentista y desde el enfoque bayesiano. El interés de este modelo *joint* reside en los coeficientes de asociación estimados para cada transición o causa específica. La respuesta longitudinal logarítmica de TRECs se asocia de manera significativa tanto con el riesgo de recaída como de muerte sin recaída,

consecuente con la asociación que se había estimado que existía con el desenlace clínico compuesto PFS.

Proceso longitudinal				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
<i>Tipo Donante</i>				
Donante DNE	Cat.ref	-	-	-
Donante HAPLO	-0.925	0.007	0.074	-1.841, 0.070
Donante HLAid	0.024	0.008	0.924	-0.462, 0.543
Donante TSCU	-1.0365	0.009	<0.001	-1.570, -0.515
Log(1+TRECs_preTPH)	0.053	0.001	0.176	-0.026, 0.133
Tiempo	0.007	0.000	<0.001	0.004, 0.009

Tabla 23. Resultados del proceso longitudinal estimado mediante la metodología *joint* para log(1+TRECs)

Proceso multiestado						
Transición	<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP HR 95%</i>
	log(1+TRECs_preTPH)	-0.097	0.907	0.002	0.046	0.825, 0.998
TRANS. 1 TPH-RECAIDA	Assoct log(1+TRECs)	-0.417	0.659	0.007	0.010	0.438, 0.906
TRANS. 2 TPH-EXITUS	Assoct log(1+TRECs)	-0.583	0.558	0.005	<0.001	0.378, 0.763

Tabla 24. Resultados del proceso multiestado estimado mediante la metodología *joint* para la respuesta longitudinal log(1+TRECs) y el tiempo hasta cada causa específica.

Para la transición 1, trasplante-recaída, se estima un HR a posteriori de 0.66 [IP95%: 0.43, 0.91], indicativo de que, para dos pacientes, el aumento de una unidad en el log(1+TRECs) conlleva una reducción del riesgo del 34% [9.4%; 56.2%] sobre la recaída a los 24 meses del trasplante, ajustado por el tipo de donante, niveles transformados pre trasplante de TRECs y medida post trasplante.

Para la transición 2, trasplante-exitus, la reducción del riesgo se estima aún más marcada por cada unidad de log(1+TRECs) relativa que incrementa, estimándose un HR a posteriori de 0.56 en un IP95% más preciso [0.38, 0.76] que el asociado al HR para la transición 1. El aumento relativo de una unidad del log(1+TRECs) tras el trasplante para dos pacientes resulta en una reducción del 44% [24%; 62%] del riesgo sobre fallecimiento sin recaída, ajustado por las mismas covariables mencionadas en el párrafo anterior.

Modelo joint multiestado para KRECs

Modelo mixto longitudinal

```
Multilong_K<-mvgllmer (list (log (1+KRECs) ~ Patología_hematológica +
log (1+KRECs preTPH) + Tiempo + (Tiempo | id), data = base_sm_IT, families = list (gaussian))
```

Modelo multiestado

```
cox_msK <- coxph(Surv(Tstart, Tstop, status) ~ log (1+KRECs preTPH) +strata(trans)+
cluster(id), data = msk,x=TRUE,model = TRUE)
```

Modelo Joint multiestado

```
interacts2 <- list("log(1+KRECs) " = ~ strata(trans) - 1)
```

```
JM_msK <- mvJointModelBayes (Multi_longK, cox_msK, timeVar = "Tiempo",
multiState = TRUE, Interactions = interacts2, data_MultiState = msK, idVar_MultiState = "id",
control = list(equal.strata.knots = TRUE, equal.strata.bound.knots = TRUE))
```

Modelo *joint* multiestado para la respuesta longitudinal log(1+KRECs) y el tiempo hasta cada causa específica

Proceso longitudinal				
<i>Variables</i>	<i>Coef.</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP 95%</i>
Hemopatía previa				
Hemopatía LA/SMD	Cat.ref.			
Hemopatía Aplasia	0.491	0.061	0.724	-3.659, 3.812
Hemopatía MM	-0.325	0.259	0.634	-1.558, 0.706
Otras hemopatías	-0.937	0.045	0.204	-2.268, 0.700
Hemopatía SLP	-1.178	0.067	0.004	-2.077, -0.426
Log(1+KRECs_preTPH)	0.093	0.007	0.174	-0.035, 0.1922
Tiempo	0.013	0.001	<0.001	0.009, 0.017

Tabla 23. Resultados del proceso longitudinal estimado mediante la metodología *joint* para log(1+KRECs)

Proceso multiestado						
Transición	<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p valor</i>	<i>IP HR 95%</i>
	log(1+KRECspreTPH)	-0.038	0.962	0.045	0.410	0.874, 1.044
TRANS. 1 TPH-RECAIDA	Assoct log(1+KRECs)	0.072	1.075	0.112	0.500	0.857, 1.331
TRANS. 2 TPH-EXITUS	Assoct log(1+KRECs)	-0.587	0.556	0.220	<0.001	0.377, 0.916

Tabla 24. Resultados del proceso multiestado estimado mediante la metodología *joint* para la respuesta longitudinal log(1+KRECs) y el tiempo hasta cada causa específica.

A diferencia de los resultados obtenidos mediante el enfoque multiestado para el biomarcador TRECs, los resultados estimados del modelo conjunto para la variable respuesta longitudinal KRECs en escala logarítmica difieren para la transición entre trasplante y recaída. Para la causa específica recaída, no se estima como estadísticamente significativa la asociación entre la dinámica longitudinal de este biomarcador y el riesgo sobre recaída, con un HR a posteriori de 1.07 [IP95%: 0.86, 1.33]. Este resultado es coherente con lo observado en el análisis descriptivo (ver Figura 5B) donde se podía apreciar cómo los pacientes que recaían a los 24 meses se asociaban con valores similares, e incluso ligeramente superiores en determinadas medidas post TPH, en comparación con los pacientes censurados a los 24 meses en los que no se desarrollaba el evento. Por lo tanto, no se puede afirmar que el biomarcador KRECs tenga un valor pronóstico en la recaída a los 24 meses del trasplante HSCT.

Sin embargo, los resultados para la otra causa específica contemplada en el modelo multiestado, el evento competitivo referente al fallecimiento sin recaída, sí que indican valor pronóstico para este biomarcador, al estimarse una asociación significativa entre los niveles transformados de KRECs y el tiempo hasta dicha causa específica. Para esa segunda transición, se estima un HR a posteriori de 0.56 [IP95%: 0.377, 0.916] que indica que, dados dos pacientes, por cada unidad relativa que aumente el $\log(1+\text{KRECs})$ tras el trasplante, se reduce en un 44% [8.4%; 62.3%] el riesgo de fallecer sin recaída a los 24 meses, ajustado por la hemopatía de origen, los niveles pre trasplante de $\log(1+\text{KRECs})$ y medida post TPH.

Modelo *joint* multiestado multivariante

Siguiendo la misma dinámica que en el Epígrafe 4.5, no sólo es interesante conocer el valor pronóstico de cada biomarcador por separado para cada causa específica, también lo es determinar el valor pronóstico cuando se integra la información de ambos de forma simultánea. Para ello, se conduce el análisis siguiendo la misma metodología aplicada en los modelos univariantes, con la única diferencia de incluir en el modelo multiestado las dos medidas pre trasplante de los biomarcadores, así como en el vector que define las interacciones entre la variable respuesta longitudinal y las transiciones. El modelo mixto longitudinal integrado en el modelo *joint* es el modelo estimado mediante *mvglmer* en el Epígrafe 4.5, por lo que se muestra únicamente el resultado del proceso de supervivencia.

Modelo mixto longitudinal

```
MultiTK<-mvgmlmer (list (log (1+TRECs) ~ Tipo_donante + log (1+TRECs preTPH) +
Tiempo +(Tiempo | id), log (1+KRECs) ~ Patología_hematológica + log (1+KRECs preTPH) +
Tiempo + (Tiempo | id)), data = base_MULTI, families = list (gaussian, gaussian))
```

Modelo multiestado

```
cox_msTK <- coxph(Surv(Tstart, Tstop, status) ~ log (1+TRECs preTPH) + log (1+KRECs preTPH)
+strata(trans)+ cluster(id), data = msk,x=TRUE,model = TRUE)
```

Modelo Joint multiestado

```
interacts3 <- list("log(1+TRECs)" = ~ strata(trans) - 1,"log(1+KRECs)" = ~ strata(trans) - 1)
```

```
JMfit <- mvJointModelBayes (MultiKT, cox_multistTK, timeVar = "Tiempo",
Interactions = interacts3, multiState = TRUE, data_MultiState = mstk,
idVar_MultiState = "id",control = list(equal.strata.knots = TRUE,
equal.strata.bound.knots = TRUE))
```

Modelo *joint* multiestado multivariante para las respuestas longitudinales log(1+TRECs) y log(1+KRECs) y el tiempo hasta cada causa específica.

Proceso multiestado							
Variable respuesta longitudinal	Transición	<i>Variables</i>	<i>Coef.</i>	<i>HR</i>	<i>Std.Err</i>	<i>p val</i>	<i>IP HR 95%</i>
		log(1+TRECs_pre)	-0.094	0.910	0.050	0.074	0.825, 1.006
		log(1+KRECs_pre)	0.003	1.003	0.046	0.918	0.914, 1.091
TRECs	TRANS. 1 TPH- RECAIDA	Assoct log(1+TRECs)	-0.278	0.757	0.007	0.026	0.549, 0.972
	TRANS. 2 TPH-EXITUS	Assoct log(1+TRECs)	-0.190	0.826	0.009	0.136	0.616, 1.039
KRECs	TRANS. 1 TPH- RECAIDA	Assoct log(1+KRECs)	0.091	1.096	0.002	0.156	0.97, 1.259
	TRANS. 2 TPH-EXITUS	Assoct log(1+KRECs)	-0.253	0.776	0.012	<0.001	0.620, 0.917

Tabla 24. Resultados del proceso multiestado multivariante estimado mediante la metodología *joint* para las respuestas longitudinales log(1+TRECs) y log(1+KRECs) y el tiempo hasta cada causa específica.

El análisis multivariante estima unos resultados que siguen la línea de lo inferido en todos los análisis previos, con la excepción del resultado estimado para la asociación entre la respuesta longitudinal de TRECs y el tiempo hasta muerte sin recaída. En el análisis

univariante de riesgos competitivos, se había estimado una asociación significativa entre los TRECs y el riesgo de recaída a los 24 meses (HR: 0.56 [0.38, 0.76]), pero al ajustar las dos respuestas longitudinales, dicha asociación deja de estimarse significativa (HR: 0.83 [0.62, 1.04]), definiéndose como factor pronóstico para esta causa específica los niveles de KRECs, que explican o recogen el efecto sobre esa transición que se había estimado para el biomarcador TRECs de forma independiente.

Los resultados estimados para la respuesta longitudinal logarítmica de KRECs, por lo tanto, mantienen una tendencia similar a lo estimado en el modelo *joint* univariante, con una asociación a posteriori estadísticamente no significativa sobre la transición 1 trasplante-recaída (HR: 1.09 [0.97, 1.26]), y una asociación a posteriori estadísticamente significativa sobre la transición 2 trasplante-fallecimiento (HR: 0.78 [0.62, 0.92]). Por lo tanto, ajustando por ambos biomarcadores, los niveles transformados de KRECs poseen valor protector sobre el fallecimiento sin recaída a los 24 meses del HSCT, pudiéndose interpretar que, para unos niveles similares de TRECs, las variaciones en los niveles post trasplante de KRECs van a asociarse con un incremento o descenso del riesgo de fallecer sin recaída en este período estudiado. Dos pacientes ajustados por las covariables incluidas en el modelo multivariante (tipo de donante, hemopatía, niveles transformados pre trasplante de TRECs y KRECs, medida post TPH) y similar comportamiento longitudinal de $\log(1+\text{TRECs})$, el aumento relativo de una unidad en los niveles post TPH de $\log(1+\text{KRECs})$ se asocia con una reducción del 22.4% [8.3%; 38.1%] del riesgo de fallecer sin recaída a los 24 meses del trasplante.

Por el contrario, para la causa específica recaída, el factor pronóstico inferido en base a los resultados es la respuesta longitudinal transformada de TRECs. En este caso, la integración simultánea de ambos biomarcadores longitudinales permite inferir que, para unos mismos niveles de $\log(1+\text{KRECs})$, los niveles post TPH de $\log(1+\text{TRECs})$ van a asociarse de forma significativa con el tiempo hasta recaída a los 24 meses del trasplante (HR: 0.76 [0.55, 0.97]). Dos pacientes ajustados por las mismas covariables mencionadas en el párrafo previo, y tendencia longitudinal similar de $\log(1+\text{KRECs})$ post trasplante, el incremento de una unidad en los niveles de $\log(1+\text{TRECs})$ implica una reducción del 24.3 % [2.8%; 45.1%] del riesgo de recaída a los 24 meses del trasplante.

Retomando las estimaciones obtenidas en el modelo *joint* multivariante para PFS, atendiendo a estos resultados se evidencia que ambos biomarcadores aportan información

similar para el riesgo sobre cada evento, implicando que, al definir un desenlace clínico con ambos eventos específicos incluidos como PFS, incremente el error estándar y se obtengan intervalos de probabilidad tan amplios como los estimados en ese modelo *joint* multivariante para PFS.

5 FORTALEZAS Y LIMITACIONES

La principal fortaleza del trabajo es el enfoque de los modelos *joint* optado para abordar el problema clínico. Las características del problema planteado encajaban a la perfección con los requisitos de esta metodología, centrando el interés en dos variables endógenas como son los TRECs y KRECs, con una serie de medidas repetidas, y con el objetivo principal de determinar su asociación con desenlaces clásicos del análisis de supervivencia. La aplicación de esta metodología ha permitido modelizar de una forma más ajustada un proceso longitudinal endógeno con un proceso de tiempo hasta evento, reduciendo los potenciales sesgos derivados de aplicar otras metodologías como la extensión del modelo de Cox para variables tiempo-dependientes.

El enfoque para el análisis de riesgos competitivos mediante los modelos multiestado sin estados intermedios también constituye una fortaleza del estudio, ya que se ha podido estimar la asociación entre las dinámicas longitudinales de TRECs y KRECs con los riesgos de cada evento competitivo, y además permite que en futuros estudios se pueda analizar la asociación de éstas con el tiempo hasta transiciones intermedias.

En cuanto a las limitaciones, mencionar que se ha trabajado con una transformación $\log(1+\text{biomarcador})$ para reducir la elevada dispersión de los valores en cada variable y obtener resultados más consistentes para los distintos objetivos planteados, a excepción del análisis multivariante para PFS como ya se ha mencionado. Se han considerado otras distribuciones que se adecuaban más a la distribución óptima para ambos biomarcadores, como una distribución *gamma*, tratando de estimar los modelos *joint* para esta familia de distribuciones. Si bien se pudieron estimar modelos longitudinales univariantes asumiendo una distribución *gamma* mediante un método de cuasi-verosimilitud penalizada (PQL), y obtener estimaciones de la asociación entre la respuesta longitudinal y el tiempo hasta PFS similares a los mostrados en el presente estudio, por limitaciones del software no se pudieron elaborar modelos multivariantes ni tampoco abordar el análisis de riesgos competitivos mediante modelos multiestado sin estados intermedios. Además, comparando los modelos asumiendo una distribución normal y una *gamma*, se observó que se ajustaban mejor los modelos asumiendo un comportamiento gaussiano.

Los resultados obtenidos se ajustan a lo esperado y observado en la literatura científica. No obstante, los resultados de los modelo multivariante, especialmente para PFS, han de ser interpretados con cautela, exigiéndose para el futuro un análisis más minucioso en otra u otras cohortes, para evaluar en qué medida se reproducen los resultados obtenidos,

y que permita asegurar la relación existente entre los biomarcadores TRECs y KRECs y su impacto conjunto sobre PFS, NRM y recaída tras un HSCT.

6 CONCLUSIONES

Atendiendo a los resultados obtenidos, se pueden extraer una serie de conclusiones de los análisis realizados:

- Se ha observado la existencia de una asociación entre cada uno de los biomarcadores TRECs y KRECs con el tiempo libre de progresión de enfermedad a los 24 meses del trasplante alogénico, ejerciendo ambos como factores protectores sobre el riesgo de evento. Este resultado sería coherente con el hecho de que, tras el “reinicio” del sistema inmune al trasplante, una rápida y eficaz reconstitución inmune adaptativa tanto de la respuesta mediada por linfocitos T, como por linfocitos B, se asocia con un mejor pronóstico a los 2 años del trasplante.
- Al integrar de forma multivariante la información de ambos biomarcadores, únicamente la respuesta longitudinal de TRECs se asocia de manera protectora con el riesgo de PFS a los 2 años del trasplante alogénico, recogiendo este biomarcador la información aportada por los KRECs para este desenlace clínico. No obstante, estos resultados han de ser interpretados con cautela debido a la magnitud de la estimación obtenida, siendo necesario ver si los resultados se reproducen en otras cohortes de pacientes.
- Se ha constatado la asociación protectora del biomarcador TRECs con cada uno de los eventos competitivos, tanto recaída como NRM, mientras que el biomarcador KRECs se asocia únicamente con el tiempo hasta fallecimiento sin recaída, ejerciendo también como factor protector. Al integrar ambos biomarcadores de forma multivariante, la tendencia de los TRECs mantiene la asociación protectora con el riesgo de recaída a los 2 años del HSCT, pero deja de asociarse con el riesgo de NRM, quedando recogido su efecto en el comportamiento de los KRECs, que continúan asociándose, con valor protector, con el tiempo hasta fallecimiento sin recaída a los 2 años del HSCT.
- La metodología de los modelos *joint* constituye una alternativa óptima para dar respuesta a problemas similares, y habituales, en la práctica clínica, especialmente

de trasplantes. En este sentido, definir mediante un método robusto el valor pronóstico de biomarcadores endógenos puede aportar información fundamental a la hora de establecer una estrategia terapéutica tras un trasplante de estas características, mejorando tanto la capacidad de actuación clínica como, en consecuencia, el pronóstico vital del paciente trasplantado.

7 BIBLIOGRAFÍA

- ¹ Twisk JWR. Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide. Cambridge: Cambridge University Press; 2003.
- ² Cox, D.R. The statistical analysis of dependencies in point processes. In *Symposium on Point Processes* (P.A.W. Lewis, ed.). New York: Wiley;1972
- ³ Laird N, y Ware J. Random-effects model for longitudinal data. *Biometrics* 1982;1:963-74
- ⁴ Self S, Pawitan Y. Modelling a marker of disease progression and onset of disease. En: *AIDS epidemiology: methodological issues*. Boston: Birkhaser, 1992.
- ⁵ Kalbfleisch, JD, Prentice R. The Statistical Analysis of Failure Time Data. New York.: John Wiley & Sons, 2002.
- ⁶ Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004;14:809-34
- ⁷ Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R. CRC Press; 2012.
- ⁸ Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical J [Internet]*. 2011 Sep 1;53(5):750–63.
- ⁹ Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study.» *Annals of Statistics* 1982;10:1100-20.
- ¹⁰ Kleinbaum DG, Klein M. Survival Analysis. A Self-Learning Text. Springer, 3ª edic. Nueva York: Springer, 2012.
- ¹¹ Faucett, C., Thomas, D.: Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Stat. Med.* 1996; **15**, 1663–1685
- ¹² Wulfsohn, M., Tsiatis, A.: A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; 53, 330–339
- ¹³ Ferrer L, Rondeau V, Dignam J, Pickles T, Jacqmin-Gadda H, Proust-Lima C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Stat Med.* 2016 ;35 , 3933-3948.
- ¹⁴ Rizopoulos D. The R package JMBayes for fitting joint models for longitudinal and time-toevent datausing MCMC. *Journal of Statistical Software* 2016; 72:1–45.
- ¹⁵ Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68 :394-424.
- ¹⁶ Copelan EA. Hematopoietic stem-cell transplantation. *N Engl j med* 2006; 354:1813–1826
- ¹⁷ Aiuti A, Scala S, Chabannon C. Biological Properties of HSC: Scientific Basis for HSCT. In: Carreras E, Dufour C, Mohty M, Kröger N, editors. *The EBMT Handbook: Hematopoietic Stem Cell Transplantation and Cellular Therapies [Internet]*. Cham: Springer International Publishing; 2019. p. 49–56.
- ¹⁸ Minculescu, L., and H. Sengeløv. "The role of gamma delta T cells in haematopoietic stem cell transplantation." *Scandinavian journal of immunology* 2015; 81.6 :459-468.

- ¹⁹ Livak F., Schatz D.G. T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol. Cell. Biol.* 1996;16(2):609
- ²⁰ Douek D.C., McFarland R.D., Keiser P.H., Gage E.A., Massey J.M., Haynes B.F. Changes in thymic function with age and during the treatment of HIV infection. *Nature.* 1998;396(6712):690
- ²¹ J.J. van Dongen, A.W. Langerak, M. Bruggemann, P.A. Evans, M. Hummel, F.L. Lavender, E. Delabesse, F. Davi, E. Schuurin, R. Garcia-Sanz, J.H. van Krieken, J. Droese, D. Gonzalez, C. Bastard, H.E. White, M. Spaargaren, M. Gonzalez, A. Parreira, J.L. Smith, G.J. Morgan, M. Kneba, E.A. Macintyre, Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936, *Leukemia* 17 (2003) 2257–2317
- ²² Sottini A, Ghidini C, Zanotti C, Chiarini M, Caimi L, Lanfranchi A, et al. Simultaneous quantification of recent thymic T-cell and bone marrow B-cell emigrants in patients with primary immunodeficiency undergone to stem cell transplantation. *Clin Immunol.* 2010;136(2):217–27.
- ²³ M.C. van Zelm, T. Szczepanski, M. van der Burg, J.J.M. van Dongen, Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion, *J. Exp. Med.* 2007; 204: 645–655.
- ²⁴ Mensen A, Jöhrens K, Anagnostopoulos I, Demski S, Oey M, Stroux A, et al. Bone marrow T-cell infiltration during acute GVHD is associated with delayed B-cell recovery and function after HSCT. *Blood.* 2014 Aug;124(6):963–72.
- ²⁵ Gaballa A, Sundin M, Stikvoort A, Abumaree M, Uzunel M, Sairafi D, et al. T Cell Receptor Excision Circle (TREC) Monitoring after Allogeneic Stem Cell Transplantation; a Predictive Marker for Complications and Clinical Outcome. *Int J Mol Sci.* 2016 Oct;17(10).
- ²⁶ Douek DC, Vescio RA, Betts MR, Brenchley JM, Hill BJ, Zhang L, et al. Assessment of thymic output in adults after haematopoietic stem-cell transplantation and prediction of T-cell reconstitution. *Lancet (London, England).* 2000 May;355(9218):1875–81.
- ²⁷ Atschekzei F, Ahmad F, Witte T, Jacobs R, Schmidt RE. Limitation of Simultaneous Analysis of T-Cell Receptor and κ -Deleting Recombination Excision Circles Based on Multiplex Real-Time Polymerase Chain Reaction in Common Variable Immunodeficiency Patients. *Int Arch Allergy Immunol.* 2016;171(2):136–40.
- ²⁸ Olbrich P, de Felipe B, Delgado-Pecellin C, Rodero R, Rojas P, Aguayo J, et al. Primer estudio piloto en España sobre el cribado neonatal de las inmunodeficiencias primarias: *An Pediatrfa.* 2014;81(5):310–7
- ²⁹ Verbeke, G. and Molenberghs, G. *Linear Mixed Models for Longitudinal Data*, New York: Springer. 2000.
- ³⁰ Díaz FJ. Introducción a los estudios de cohorte en epidemiología y al análisis de supervivencia. *MedUNAB.* 2018 May ;8(1):43
- ³¹ Brun González LP, Salazar Uribe JC. Efecto de la censura informativa sobre la potencia de algunas pruebas tipo Log-Rank .Vol. 7, *Ciencia en Desarrollo . scieloco* ; 2016. p. 45–53.
- ³² Prentice RL, Kalbfleisch JD, Peterson A V, Flournoy N, Farewell VT, Breslow NE. The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics [Internet].* 1978 Jun 17;34(4):541–54.

- ³³ Aalen OO. Further results on the non-parametric linear regression model in survival analysis. *Stat Med*. 1993 Sep 15;12(17):1569–88.
- ³⁴ Fine, J. y Gray, R. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*. 1999; 94(446), 496-509.
- ³⁵ Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. *Stat Methods Med Res*. 2002 Apr;11(2):203–15.
- ³⁶ Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000 Dec;1(4):465–80.
- ³⁷ Wang Y, Taylor JMG. Jointly Modeling Longitudinal and Event Time Data With Application to Acquired Immunodeficiency Syndrome. *J Am Stat Assoc*. 2001 Sep 1;96(455):895–905.
- ³⁸ Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*. 2006;62(2):432–445.
- ³⁹ Mauff K, Steyerberg E, Kardys I, Boersma E, Rizopoulos D. Joint Models with Multiple Longitudinal Outcomes and a Time-to-Event Outcome. 2018.
- ⁴⁰ Proust-Lima C, Joly P, Dartigues J-F, Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach, *Computational Statistics and Data Analysis*, 2009, vol. 53 (pg. 1142-1154).
- ⁴² de Wreede LC, Fiocco M, Putter H (2010). “The mstate Package for Estimation and Prediction in Non- and Semi-Parametric Multi-State and Competing Risks Models.” *Computer Methods and Programs in Biomedicine*, 99, 261–274b

ANEXO

Código para el modelo longitudinal de efectos mixtos.

Se muestra el código aplicado para estimar el modelo efectos mixtos que compone proceso longitudinal del modelo joint, incluyendo como estructura de efectos aleatorios tanto término independiente aleatorio (*id*) como pendiente aleatoria (*tiempo_n*).

#Librería utilizada

```
library(nlme)
```

#Modelos univariantes. Sólo para TRECs, siendo similar para KRECs modificando la variable respuesta (KRECs_trans) e incluyendo un componente de control para la convergencia (control = lmeControl(opt = "optim"))).

#Sexo

```
l1<-lme(TRECs_trans ~ SEXO_n,random = ~tiempo_n|id,base_sm_IT)
```

#Tipo de donante

```
l2<-lme(TRECs_trans ~ TIPO_TPH_n,random = ~tiempo_n|id,base_sm_IT)  
anova(l2)
```

#Edad al trasplante

```
l3<-lme(TRECs_trans ~ EDAD_TPH,random = ~tiempo_n|id,base_sm_IT)
```

#Acondicionamiento previo

```
l4<-lme(TRECs_trans ~ ACOND_n,random = ~tiempo_n|id,base_sm_IT)
```

#log (1+TRECs preTPH)

```
l7<-lme(TRECs_trans ~ TRECspre_trans,random = ~tiempo_n|id,base_sm_IT)
```

#EICH

```
l8<-lme(TRECs_trans ~ EICH_SINO_n,random = ~tiempo_n|id,base_sm_IT)
```

#Patología hematológica

```
l9<-lme(TRECs_trans ~ PAT_HEM_n,random = ~tiempo_n|id,base_sm_IT)  
anova(l9)
```

#Tiempo

```
l10<-lme(TRECs_trans ~ tiempo_n,random = ~tiempo_n|id,base_sm_IT)
```

#Modelo multivariable definitivo

TRECs

```
t1_def<-lme (TRECs_trans ~ TIPO_TPH_n +tiempo_n+TRECspre_trans,  
random = ~ tiempo_n|id,data = base_sm_IT)
```

Código para el modelo de COX de riesgos proporcionales para PFS

Se muestra el código aplicado para estimar el modelo de regresión de COX que compone el proceso de supervivencia del modelo joint. El desenlace clínico es el tiempo libre de progresión PFS.

#Libreria utilizada

```
library(survival)
```

#Modelos univariables

#Sexo

```
cox1<-coxph (Surv (tiseq, PFS) ~ SEXO_n,baseimpwide_TREC)
```

#Tipo de donante

```
cox3<-coxph (Surv (tiseq, PFS) ~ TIPO_TPH_n,baseimpwide_TREC)  
anova(cox3)
```

#Edad al trasplante

```
cox2<-coxph (Surv (tiseq, PFS) ~ EDAD_TPH,baseimpwide_TREC)
```

#Acondicionamiento previo

```
cox4<-coxph (Surv (tiseq, PFS) ~ ACOND_n,baseimpwide_TREC)
```

#log (1+TRECs preTPH)

```
cox7<-coxph (Surv (tiseq, PFS) ~ TRECspre_trans,baseimpwide_TREC)
```

#EICH

```
cox5<-coxph (Surv (tiseq, PFS) ~ EICH_SINO_n,baseimpwide_TREC)
```

#Patología hematológica

```
cox6<-coxph (Surv (tiseq, PFS) ~ PAT_HEM_n,baseimpwide_TREC)  
anova(cox6)
```

#Modelo definitivo

TRECs

```
cox_def<-coxph (Surv (tiseq,PFS) ~ TRECspre_trans , baseimpwide_TREC)
```


Código para el modelo *joint* univariante

Se muestra el código aplicado para estimar el modelo *joint* univariante para cada biomarcador para el evento PFS. A partir de aquí se utilizó la misma semilla para todos los modelos *joint*, tanto univariante como multivariante.

#Libreria utilizada

```
library(JMBayes)
```

#Semilla

```
set.seed(4062020)
```

##Modelo Joint univariante

```
joint_T<-jointModelBayes(t1_DEF,coxt_def,timeVar = "tiempo_n")
```

Código para el modelo *joint* multivariante

Se muestra el código aplicado para estimar el modelo joint multivariante para TRECs y KRECs de forma conjunta para el evento *PFS*

#Libreria utilizada

```
library(JMBayes)
```

```
library(survival)
```

#Modelo mixto multivariante

```
MultiKT<-mvglmer(list(
TRECs_trans~TIPO_TPH_n+TRECspre_trans+tiempo_n+(tiempo_n|id),
KRECs_trans~PAT_HEM_n+KRECspre_trans+tiempo_n+(tiempo_n|id)),
data=base_MULTI,families=list(gaussian,gaussian))
```

#Modelo de regresión de Cox con ambas medidas preTPH.

```
cox_multiKT<-coxph (Surv (tiseg,PFS)~TRECspre_trans+
KRECspre_trans,data = base_multicox,x=TRUE,model=TRUE)
```

#Modelo joint multivariante.

```
JBmulti<-mvJointModelBayes(MultiKT,cox_multiKT, timeVar = "tiempo_n")
```

Código para el modelo *joint* univariante de riesgos competitivos

Se muestra el código aplicado para estimar el modelo *joint* univariante para el análisis de riesgos competitivos para cada biomarcador. Se aborda mediante un enfoque multiestado sin eventos intermedios. El código es para obtener el modelo univariante para TRECs, siendo similar la metodología para KRECs.

#Librería utilizada

```
library(msprep)
```

```
library(JMBayes)
```

#Construcción de base de datos en formato multiestado

#Creamos matriz de transición

```
tmat1<- transMat(x = list(c(2,3), c(),c()),  
                 names = c("Trasp", "Recaída", "Muerte"))
```

#Covariable a mantener en la base de datos.

```
keep1<-data.frame(TRECspre_trans=baseimpwide_TREC$TRECspre_trans)
```

#Se construye la base de datos. rtime: tiempo hasta recaída; etime: tiempo hasta exitus sin recaída.

```
mst<-msprep(time = c(NA,"rtime","etime"),  
            status = c(NA, "recaída", "exitus"),  
            data = baseimpwide_TREC,  
            keep = keep1,  
            trans = tmat1)
```

#Modelo mixto univariante

```
Multilong_T<-mvglmer (list (TRECs_trans~TIPO_TPH_n+TRECspre_trans+  
tiempo_n+ (tiempo_n|id)),data=base_sm_IT, families=list(gaussian))
```

#Modelo de cox multiestado

```
cox_msT <- coxph (Surv (Tstart, Tstop, status) ~ TRECspre_trans+ strata (trans)+cluster (id),  
data = mst,x=TRUE,model = TRUE)
```

#Componente para estimar la asociación de la respuesta longitudinal con cada transición contemplada

```
interacts <- list("TRECs_trans" = ~ strata(trans) - 1)
```

#Modelo joint univariante para riesgos competitivos.

```
JM_msT <- mvJointModelBayes(Multi_longT, cox_msT,  
                             timeVar = "tiempo_n",  
                             multiState = TRUE,  
                             Interactions = interacts,  
                             data_MultiState = mst,  
                             idVar_MultiState = "id",  
                             control = list( equal.strata.knots= TRUE, equal.strata.bound.knots = TRUE))
```

Código para el modelo *joint* multivariante de riesgos competitivos

Se muestra el código aplicado para estimar el modelo *joint* multivariante para el análisis de riesgos competitivos para ambos biomarcadores.

#Libreria utilizada

```
library(msprep)  
library(JMBayes)
```

#Se modifican las variables a mantener

```
keep3<-data.frame(TRECspre_trans=base_mstate$TRECspre_trans,  
KRECspre_trans=base_mstate$KRECspre_trans)
```

#Modelo mixto multivariante

```
MultiKT<-mvglmr(list(  
TRECs_trans~TIPO_TPH_n+TRECspre_trans+tiempo_n+(tiempo_n|id),  
KRECs_trans~PAT_HEM_n+KRECspre_trans+tiempo_n+(tiempo_n|id)),  
data=base_MULTI,families=list(gaussian,gaussian))
```

#Modelo de cox multiestado

```
cox_multistTK <- coxph(Surv(Tstart, Tstop, status) ~ TRECspre_trans+  
KRECspre_trans +strata(trans)+cluster(id), data = mst,x=TRUE,model = TRUE)
```

#Componente de interacción

```
interacts3 <- list("TRECs_trans" = ~ strata(trans) - 1,"KRECs_trans" = ~ strata(trans) - 1)
```

#Modelo multivariante JOINT

```
JMfit <- mvJointModelBayes(MultiKT, cox_multistTK,  
                           timeVar = "tiempo_n",  
                           Interactions = interacts3,  
                           multiState = TRUE,  
                           data_MultiState = mstk,  
                           idVar_MultiState = "id",  
                           control = list(equal.strata.knots = TRUE, equal.strata.bound.knots = TRUE))
```

Código para las figuras incluidas en el análisis descriptivo

A) Figura 2 y 3 → Gráficos de densidad TRECs y KRECs en crudo (A) y transformadas (B)

#Librerías utilizadas

```
library(ggplot2)  
library(dslabs)  
library(survminer)  
library(survival)  
library(lattice)  
library(dplyr)
```

#TRECs crudo. Similar para KRECs cambiando TRECs por KRECs

```
ds_theme_set()
```

```
base_imputada%>%filter(!is.na (TRECs))%>%ggplot()+  
geom_density (aes (TRECs, ..count.., fill=as.factor(tiempo)),  
alpha=0.2, bw=0.75, col="blue", show.legend = FALSE)+ facet_grid(tiempo ~.)
```

#log(1+TRECs). Similar para log(1+KRECs) cambiando TRECs_trans por KRECs_trans

```
base_imputada%>%filter (!is.na (TRECs_trans))%>%ggplot()+  
geom_density (aes (TRECs_trans, ..count..,fill=as.factor(tiempo)),  
alpha=0.2, bw=0.75, col="blue", show.legend = FALSE)+ facet_grid(tiempo ~.)
```

B) Figura 4 → Gráfica de la curva de supervivencia de Kaplan-Meier para PFS:

#Función de supervivencia para PFS

```
KM_PFS<-survfit(Surv(tiseg,PFS)~1,baseimp_wideTREC)
```

#Curva de supervivencia para PFS

```
ggsurvplot(KM_PFS,title="Curva de supervivencia PFS 24m", break.time.by=60,risk.table =  
T,palette = c("navyblue"), ggtheme = theme_pubr())
```

- C) Figura 5 → Gráficos de cajas para el biomarcador estratificado por los 3 desenlaces clínicos PFS, Recaída y NRM. Se muestra el código para elaborar los gráficos para TRECs, siendo similar para KRECs.

#PFS

```
bwplot(TRECs_trans ~ tiempo|PFStotal, data=base_imputada,
,aspect = 2, ylab="log(1+TRECs)", xlab="tiempo",main="PFS 24m")
```

#Recaída

```
bwplot(TRECs_trans ~ tiempo|recaidatotal, data=base_imputada ,
,aspect = 2, ylab="log(1+TRECs)", xlab="tiempo",main="Recaída 24m")
```

#NRM

```
bwplot(TRECs_trans ~ tiempo|NRMtotal, data=base_imputada ,
,aspect = 2, ylab="log(1+TRECs)", xlab="tiempo",main="NRM 24m")
```

- D) Figura 6 → Gráficos de perfil para tendencia longitudinal del biomarcador post TPH.

#5 pacientes seleccionados al azar

```
id2<-sample(base_imputada$id,5,replace = FALSE)
```

#Gráfico de perfil para TRECs. Similar para KRECs, modificando TRECs_trans por KRECs trans.

```
base_sm_IT %>% filter (id%in%id2)%>%ggplot (aes (ts,TRECs_trans))+geom_line()+
facet_wrap(~id) + xlab("tiempo") + ylab ("log(1+TRECs)") + ggtitle("Tendencia individual")
```